

图灵数学·统计学丛书



Probability and Statistics

概率统计

（第3版）

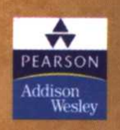
［美］ Morris H. DeGroot
Mark J. Schervish 著

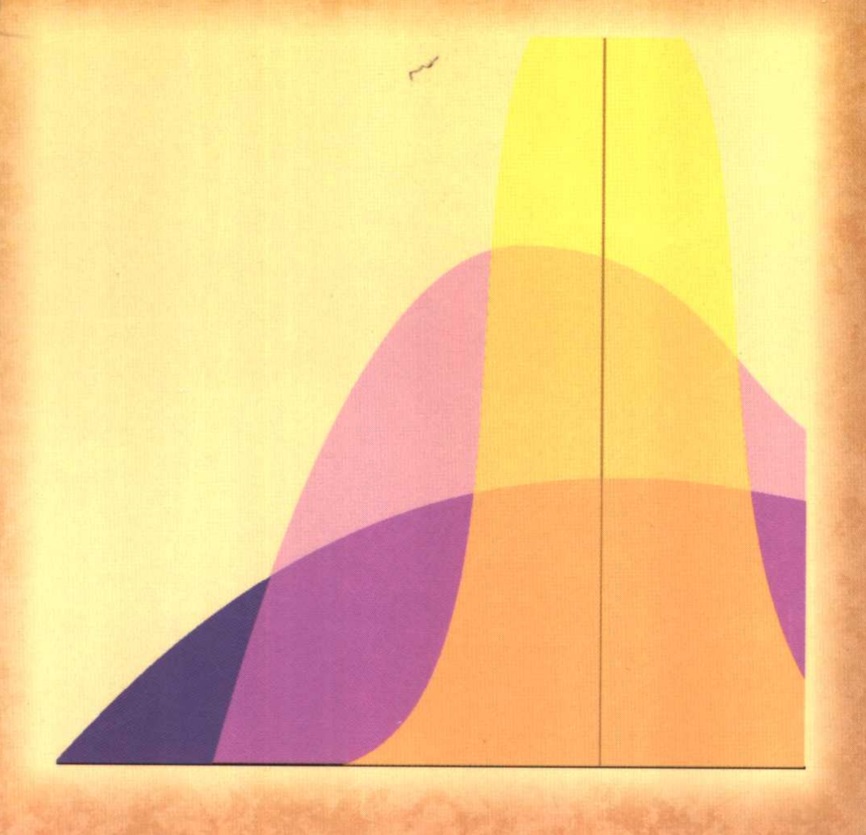
叶中行 王蓉华 徐晓岭 译

人民邮电出版社
北 京



图灵数学·统计学丛书







Probability and Statistics

概率统计

(第3版)

[美] Morris H. DeGroot
Mark J. Schervish 著
叶中行 王蓉华 徐晓岭 译



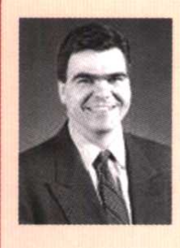
概率统计（第3版）

Probability and Statistics

这是一本堪称典范的概率论与数理统计教材，被很多知名大学采用，包括卡内基梅隆大学、哈佛大学、麻省理工学院、华盛顿大学、杜克大学、加利福尼亚大学洛杉矶分校等。

概率论历来以抽象难学著称，而本书作者以大量形象的例题阐释理论和证明，使内容深入浅出。本书例题涉及面广泛，除通过一些著名的例题来解释概率中的基本概念之外，又加入了一些新的例题，描述了概率论在遗传学、排队论和计算金融学中的应用。还将目前研究前沿的一些问题深入浅出地融入教材。

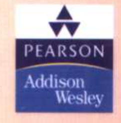
全书内容丰富完整，除了经典的概率理论之外，还有一大部分介绍估计方法（极大似然估计、贝叶斯估计、最小二乘法），并讨论了统计检验和其他非参数方法以及随机模拟等。



Morris H. DeGroot (1931–1989) 世界知名统计学家。生前是国际统计学会、美国统计学会、数理统计学会、计量经济学会会士。卡内基梅隆大学教授，1966年创办该校统计系。DeGroot在学术上非常活跃，著述颇丰。为纪念他的著作对统计教学的贡献，国际贝叶斯分析学会特别设立了DeGroot奖表彰优秀统计学著作。



Mark J. Schervish 世界知名统计学家，美国统计学会、数理统计学会会士。卡内基梅隆大学统计系系主任。除本书外，还著有 *Theory of Statistics* 和 *Rethinking the Foundations of Statistics*。



www.PearsonEd.com

本书相关信息请访问：
图灵网站 http://www.turingbook.com
读者/作者热线：(010) 88593802
反馈/投稿/推荐信箱：contact@turingbook.com

分类建议 概率论

人民邮电出版社网址 www.ptpress.com.cn



ISBN 978-7-115-13913-9/O1

定价：59.00 元

译者简介

叶中行 美国康奈尔(Cornell)大学博士，现任上海交通大学理学院副院长，现代金融研究中心副主任，教授，博士生导师，中国工程概率统计学会理事长. 其研究领域包括应用概率统计、金融数学、信息理论和信息处理、智能计算等. 出版专著、译著和教材6本，在国内外学术期刊和国际学术会议发表论文近百篇.

王蓉华 上海师范大学数理信息学院副教授，硕士生导师，应用数学与统计系系主任，统计专业学科带头人. 研究领域包括应用统计、统计教育等. 在国内外学术期刊和国际学术会议发表论文40余篇.

徐晓岭 上海对外贸易学院国际经贸学院教授，统计学博士，硕士生导师. 中国运筹学会可靠性数学分会理事，上海现场统计研究会理事. 研究领域包括应用统计、商务统计等. 在国内外学术期刊和国际学术会议发表论文70余篇.

译 者 序

本书是国际上一本畅销的大学本科生用的“概率论和数理统计”教材，原著作者的序言已经详细介绍了本书的特点．全书内容丰富完整，特色明显，语言生动流畅，例题涉及面广泛，除了通过一些著名的例子来解释概率中的基本概念之外，又加入了一些新的例子，描述了概率论在遗传学、排队论和计算金融学中的应用．将目前研究前沿的一些问题深入浅出地融入了教材，对提高大学本科生学习本课程的兴趣、了解应用背景很有帮助．

根据原作者和人民邮电出版社编辑人员的建议，我们删节了原书的部分内容，其中有属于中学阶段已经学习过的比较浅显的内容，如排列组合的基本知识，也有属于超出本科生水平的比较艰深的内容，此外还删除了原附在各章之后的难度较大的补充习题．对保留部分中涉及删除部分内容的作了适当增减，使得在删除了上述内容后的这个版本仍是一个自成体系的完整版本．当然如果读者对删除部分感兴趣的话，可以参考原著．

本书具有广泛的适用性，既可作为大学各专业的本科生教材，也可以作为实务工作者的参考书．

本书的翻译工作是由上海交通大学的叶中行、上海师范大学的王蓉华、上海对外贸易学院的徐晓岭和华东理工大学的马艳梅通力合作完成的．叶中行负责协调全书的翻译工作并负责翻译前四章以及附录部分(包括习题解、名词索引等)，马艳梅翻译了第5章，王蓉华翻译了第6～9章，徐晓岭翻译了第10、11章，叶中行还对全书译稿进行了校对和润色．

上海交通大学数学系、河南大学数学系、华东理工大学商学院的部分研究生和进修教师对部分章节进行了试译，并帮助译文的计算机录入，他们是王起为、陈志娟、李灏、王英泽、王元英、胡华、左佳明、陆青、唐彦斌、符丽健、赵辉、秦长城、陈蕾、胡冰、辛欣、李俊、赵克、张志娟、高莉、杨菲菲、王丹娜、李国凤和杜利敏，在此向他们表示衷心感谢．译者感谢人民邮电出版社图灵公司的指导和支持，使本书的翻译出版成为现实．

译 者
2006 年 6 月于上海

前　　言

Morris DeGroot 已于 1989 年 11 月 2 日去世，但是他的思想并没有随他而去. 我很荣幸能够通过本书的第 3 版来继续发扬 Morris 的传统，虽然在这里做了一些改动，但我尽量保留本书的风格，尤其是 Morris 特有的叙述清晰的风格，使本书仍然卓而不群，引人入胜. 基于这个原因，我尽量使下边的前言与 Morris 本人写的第 2 版的前言风格一致. 在第 3 版中改变了那些不再正确的论述，并且对这些改动做了解释，因为我至少要再给 Morris 一次用他自己的表达方式来介绍他的书的机会.

这本书包含的内容比一年的概率统计课程稍多，对本课程数学上的要求是了解微积分的基本概念，熟悉向量和矩阵的概念和基本性质，这里假定读者事先没有概率和统计的相关知识.

编写这本书时我们充分考虑到了学生和老师，特别注意使课文中没有模糊的片断或其他的绊脚石，在合适的地方给出了定理和证明，并给出了说明性的例题. 在本书各节的末尾包含了近 1000 道习题，有些习题是前面课文中数值结果的应用，另一些意在进一步深化这些结果的思想.

前五章的内容主要讲概率知识，它可以作为一学期课程的内容来讲. 概率中的基本概念通过一些著名的例子加以解释，比如，出生问题、网球锦标赛问题、匹配问题、收藏品问题、掷骰子游戏(game of craps)等. 通过讨论统计赝品、伪随机数的应用、预测性的均值和中位数的优劣、中心极限定理的重要性、连续性的修正，展现了关于随机变量及其概率分布的标准材料. 这些章节的特色还包括马氏链、赌徒输光问题、赌博中的效用和偏好问题，这些问题都是些初等问题，但是如果课时有限，这些问题可以省略掉，而不会损害全书连贯性.

接下来的五章主要讲的是统计推断，所讲的内容总体上是现代的. 古典概型和贝叶斯统计方法是通过完整的形式展开，而不是教条地处理哪个学派的思想，我的目标是给学生提供一些已证实有用且今后依然要用到的理论和方法. 这些章节包含了关于估计、假设检验、非参数方法、多元回归和方差分析这些广泛但基本的内容. 对作为基本概念的极大似然估计、贝叶斯决策过程、无偏估计、置信区间、显著性水平等的强弱性和优缺点等作了详细讨论. 这些章节的显著特点包括：先验和后验分布、δ 方法、无偏性检验、拟合度检验、列联表和修正均值、回归线置信带、简单线性回归的贝叶斯分析和残差分析.

第 3 版中的主要改变有八种类型，把它归纳如下：

1. 新增加的第 11 章是关于模拟方面的. 它包含模拟特定的分布、重要取样、马

尔可夫链、蒙特卡罗方法和自助法. 其中特别注意对模拟近似中不肯定性估计，例子都是从书中前面的那些较难分析的问题中选出来的.

2. 此外，又加入了一些新的例子，其中许多的例子使用了已公开发表的论文中的实际数据，数据的分析不是来自于公开出版的资料，而是来自于对本书中的概念的解释. 在这本书中很多的数据都是从下面两个好的数据源处取得的: *Data and Story Library*(DASL)和 Hand et al. (1994). DASL 项目中心设在康奈尔大学，能够在互联网上找到数据集，网址是：http://lib.stat.cmu.edu/DASL/. 由 Hand 等写的书包含了对 500 多个数据组的描述和一个磁盘以免去自己动手输入. 第 2 版中的大部分例子在第 3 版中被保留，只换掉了少部分. 这些新的例子描述了概率论在遗传学、排队论和计算金融学中的应用.

3. 我们新增加了一些章节段落，讨论条件独立事件和随机变量、对数正态分布、分位数、预测和预测区间、不完全先验(improper priors)、贝叶斯检验、幂函数、M 估计、线性模型中的残差部分和简单线性回归模型中的贝叶斯分析等内容.

4. 在每个技术性章节的前面增加了简介和综述. 这些简介的段落为读者提供将要讨论的内容的提示，综述不是对所有内容的扼要重述，而是列举了最重要的思想. 一些章节内容很丰富，虽然这些内容都很有用，但有些内容对于继续学习是重要的. 并不是所有的学生在第一次阅读时就能区分轻重缓急(但愿书中已经做到这一点). 而综述的作用是帮助指明重要的主题.

5. 在应该略加总结或与课文中别处有联系的地方增加了特殊的注释，在课文中用“注释”表示出来.

6. 课文中的内容也做了一些调整，独立性放在了条件概率的后面，第 2 版中的几个小的段落以注释或习题的形式重新编写，它们包括“最优选择”、“波雷尔 - 柯尔莫戈罗夫悖论”、“中位数和其他分位数的推断”、“回归谬论”. 所有随机数字的表格讨论被模拟内容所代替. 第 2 版中关于迹和秩的检验的两节被合并成一节. 调整最多的是第 8 章. 我完全重写了关于假设检验的第一节，给出了充分的背景材料以便能够在 8.3 节中直接给出 t 检验. 中间插入的最优检验是选读内容. 这些段落中的内容尽管数学上是漂亮的，但是在现代统计分析中已经没有那么重要了. 另外扩展了幂函数的覆盖面并介绍了有偏 t 分布.

7. 我删掉了一些第 2 版的内容，除了上面提到的改动之外，我还删除了“多决策问题”章节. 我删去了那些我认为没有达到一定深度的一些讨论，如似然准则、高斯 - 马尔可夫定理和逐步回归法.

8. 从许多章节中选出了一些较有难度的内容，把它们移到章节的末尾，作为扩展研究.

如果没有足够的时间学完这本书的话，下面这些章节可以省略掉而不会严重影响后面章节的学习：2.4，7.8、8.2，9.4 和 11.3. 除了这些章节相互交叉参考外，偶而也有其他内容需要参考这些章节，但影响不大. 多数情况是第 11 章中向前参见

某章节，原因是这些选读章节涉及了一些比较难学的内容，并且模拟方法在解决那些通过分析不能解决的难题时很有用.

虽然计算机在概率和统计课程中是很有用的辅助工具，但本书前十章大部分的习题中并不需要计算机或程序设计的知识.（少数新加的关于实数和残差点习题除外.）基于这一点，本书在第11章之前的章节和计算机应用的联系并不紧密，虽然如此，我还是为那些想要学习与模拟类似的统计方法的人在前面章节中增加了许多理论和方法方面的练习，我们鼓励教师在讲课中尽可能地应用计算机，还有小的计算器有助于解某些数值习题. Morris在去世之前，开始显示出对计算机的兴趣，这与我第一次见到他时形成鲜明的对比. 我们系第一次在各教师办公室安装电脑终端后，Morris长期置之不理，后来，当他看到文字处理软件能像打字机一样令人满意地打印文字，他才开始用计算机来处理文档. Morris去世后的这些年，计算机在统计理论和方法中发挥了更大的作用，我想，如果Morris还活着，他在讲授统计理论和方法时肯定无法离开电脑了.

我要感谢在本书修订过程中很多人给予我的鼓励和帮助，其中特别感谢Marilyn DeGroot和Morris的子女给我这个修订Morris杰作的机会.

我也感激许多读者、审阅人、同事、员工和Addison-Wesley公司的人们，他们对本版本的完成给予了大量帮助和建议. 审阅人有：华盛顿大学（圣路易斯）的Brian Blank，波士顿学院的Daniel Chambers，多伦多大学的Michael Evans，宾州印第安那大学的Doug Frank，兰道夫－梅肯学院的Lyn Geisler，俄亥俄州立大学的Prem Goel，所罗马州立大学的Susan Herring，北爱荷华大学（色达瀑布城）的Syed Kirmani，杜克大学的Michael Lavine，图雷大学的John Liukkonen，东北大学的Rosa Matzkin，塞拉卡斯大学的Terry McConnell，加州大学（戴维斯）的Hans-Georg Mueller，俄亥俄州立大学的Mario Peruggia，宾州大学的HenSiong Tan，约翰－霍普金斯大学的Kenneth Troske，俄亥俄州立大学的Joseph Verducci，肯特州立大学的Mahbobeh Vezveai，杜克大学的Brani Vidakovic，东密歇根大学的Bette Warren，克莱门森大学的Calvin L. Williams. 校稿人包括：俄亥俄州立大学的Joseph Verducci，所罗马州立大学的Susan Herring，罗格斯大学的Yehuda Vardi. 我非常感谢我在卡内基－梅隆大学的同事，特别是Anthony Brockwell，Joel Greenhouse，John Lehoczky，Heidi Sestrich和Valerie Ventura.

Addison-Wesley公司和其他机构中帮助编写这本书的人有：Paul Anagnostopoulos，Cindy Cody，Deirdre Lynch，George Nichols，Joe Snowden和Anna Stillner.

如果我漏掉了某些人，很抱歉，我不是故意的. 类似的错误也不可避免地会出现在任何类似的项目中（我的意思是说我参加过的项目）. 基于这个原因，只要这本书一出版，我就在我的网页上公布这本书的信息，包括修正勘误表，我的网页是：http://www.stat.cmu.edu/~mark/. 欢迎读者把发现的错误发给我.

自从Morris于1985年10月为这本书的第2版写过序言之后，统计学领域发生了

很大的变化. 个人计算机(工作站)正在由昂贵的稀罕的玩具逐步变成重要的研究工具. 1985 年以来, 常用的统计方法已经进一步发展并相当依赖于高速计算. 然而, 一个学生可用于该领域的学习时间并没有增加, 因此我们的书不得不放弃一些内容. 近年来, 有一个在教学中增加计算方法而减少数学理论的趋势, 我不想伪称这是一本统计计算方法的课本, 但是我努力把它写好, 并使它充分地现代化以适应新一代统计人员的需要. 我希望我已经成功达到了 Morris 家人对我的期望.

图书在版编目（CIP）数据

概率统计：第3版／（美）德格鲁特，（美）舍维什著；叶中行，王蓉华，徐晓岭译.
—北京：人民邮电出版社，2007.3
（图灵数学·统计学丛书）

ISBN 978-7-115-13913-9

Ⅰ. 概…　Ⅱ. ①德…②舍…③叶…④王…⑤徐…　Ⅲ. ①概率论②数理统计　Ⅳ. 021

中国版本图书馆CIP数据核字（2006）第132103号

内容提要

本书内容包括概率论、数理统计两部分，涉及条件概率、随机变量及其分布、数学期望、特殊分布、估计、估计量的抽样分布、假设检验、分类数据与非参数方法、线性统计模型、随机模拟等. 本书知识体系结构与国内主流的概率论与数理统计教材基本一致，但内容取材及例题的安排上都比较新颖，尤其是新增加了一些非常实用而且比较先进的模拟方法. 书中最后提供了奇数号习题的解答以及索引.

本书作为概率论与数理统计的教材，适合大学本科高年级学生和研究生用作教材或参考书，也可供统计工作人员用作参考书.

图灵数学·统计学丛书

概率统计（第3版）

- ◆ 著　[美]Morris H.DeGroot　Mark J.Schervish
  译　叶中行　王蓉华　徐晓岭
  责任编辑　明永玲　李　颖
- ◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街14号
  邮编　100061　电子函件　315@ptpress.com.cn
  网址　http://www.ptpress.com.cn
  北京铭成印刷有限公司印刷
  新华书店总店北京发行所经销
- ◆ 开本：700×1000　1/16
  印张：29.5
  字数：628千字　　2007年3月第1版
  印数：1–4 000册　　2007年3月北京第1次印刷

著作权合同登记号　图字：01-2006-0309号

ISBN 978-7-115-13913-9/O1

定价：59.00元

读者服务热线：(010)88593802　印装质量热线：(010)67129223

目 录

第 1 章　概率导引

1.1　概率的发展历史

用概率来度量不确定性和可变性已经有数百年的历史了，而今概率已应用到很多的领域，例如：医药、赌博、天气预报和法律.

偶然性和不确定性的概念像文明本身一样古老，人们不得不应付天气、食物供应和环境中其他方面的不确定性，并且为减少这种不确定性及其影响而奋斗. 即便是赌博的思想也有很长的历史. 大约公元前 3500 年，在埃及等地已经出现利用骨制物体做具有偶然性的赌博，它就是掷骰子的先驱. 立方体的骰子实质上与从公元前 2000 年的古埃及坟墓中出土的骰子完全一样. 我们知道从那时起用骰子赌博已经开始流行，并在概率理论的早期发展中扮演了重要的角色.

通常大家都认为概率的数学理论是由法国数学家帕斯卡(1623—1662)和费马(1601—1665)首创的，他们成功地导出了包括掷骰子在内的一些赌博的实际概率. 他们解决的问题中有些已是 300 多年悬而未决的. 然而更早的卡尔达诺(1501—1576)和伽利略(1564—1642)已经算出掷骰子的各种排列组合的数值概率.

自从 17 世纪以来概率论得到了稳步的发展，并且被广泛地应用到很多研究领域. 现在，概率论在工程、科学和管理等方面是一个很重要的工具. 许多研究工作者正致力于发现和确立概率论新的应用方面，比如医药、气象、卫星摄影、营销、地震预测、行为学、计算机系统设计、金融、遗传学和法律. 在反托拉斯或者雇佣歧视等法律诉讼程序中，诉讼双方都会提供概率统计的计算来支持他们的案例.

1

参考文献

David(1988)、Ore(1960)、Stigler(1986)和 Todhunter(1865)论述了赌博的古代史和概率的数学理论的起源.

一些介绍概率理论的书籍，与本书所研究的主题大致相同，较好的书有：Feller(1968)；Hoel，Port and Stone(1971)；Meyer(1970)和 Olkin，Gleser and Derman(1980). 另外一些书籍阐述概率和统计的深度上与本书相当，它们是：Brunk(1975)；Devore(1999)；Fraser(1976)；Hogg and Tanis(1997)；Kempthorne and Folks(1971)；Larsen and Marx(2001)；Larson(1974)；Lindgren(1976)；Miller and Miller(1999)；Mood，Graybill and Boes(1974)；Rice(1995) and Wackerly，Mendenhall and Schaeffer(1996).

1.2 对概率的几种解释

本节描述了概率的三种常用的运算上的解释. 尽管这些解释看起来似乎是不尽相同的, 但幸运的是无论你选择哪种解释, 同样都能很好地运用到概率的计算中去(本书前五章的主要内容).

除了概率论的许多形式的应用之外, 概率的概念已经渗透到我们的日常生活和交往中. 我们经常听到或使用这样的说法: "明天下午可能会下雨"; "飞机可能会晚点到达" 或者 "他能够有机会来参加我们今天的晚宴真好." 每种表达都是基于概率的概念, 或者说, 某个特殊事件发生的可能性.

尽管事实上概率的概念在我们的生活中是如此的普遍和自然, 可是却没有为所有统计人员、哲学家和其他的权威人士所接受的对概率这一术语的唯一科学解释. 这些年来, 一些权威人士提出的每个概率概念都受到其他人的批评. 事实上, 概率真实的含义已经成为一个非常有争议的课题, 并且涉及目前关于统计基础的哲学讨论. 这里将讨论对概率的三种不同的解释, 每种解释对于将概率论应用到实际问题中都是非常有用的.

1.2.1 概率的频率解释

在很多问题中, 对一个过程中某个特定结果的概率可被理解为该过程在相同条件下重复很多次而得到这个结果的相对频率(relative frequency). 例如, 认为一枚硬币投掷一次得到正面的概率是1/2, 是因为它在相同条件下投掷很多次得到正面的相对频率大约是1/2. 换句话说, 假定投掷得到正面的比例大约是1/2.

2

当然, 这个例子中提到的情况太模糊, 不能作为概率的科学定义的一个根据. 首先, 提到投掷硬币 "很多次", 但是对于具体的次数没有明确的界定. 其次, 规定每次必须在 "相同条件下" 投掷硬币, 但是并未确切地描述这些条件. 每次投掷硬币的条件不一定完全相同, 否则会出现相同的结果, 即全是正面或全是负面的. 事实上, 一个熟练的人可将硬币投到空中, 并且几乎每次都能以得到正面的方式抓住它. 因此, 完全控制投掷是不可能的, 必须有一些 "随机" 的特征.

而且, 得到正面的相对频率被说明为 "近似是1/2", 但是对于偏离1/2的程度没有明确的界限. 如果投掷一枚硬币1 000 000次, 我们并不期望恰好得到500 000次正面, 事实上, 如果恰好得到500 000次正面, 我们会非常惊奇. 另一方面, 我们也不期望得到正面的次数远离500 000次. 希望能够对于出现正面的不同的可能次数作一个准确的描述, 但这些可能性当然依赖于我们将要定义的概率的概念.

概率的频率解释的另一个缺点在于, 至少在原则上, 只适用于对其中一个过程能大量重复的问题. 很多重要的问题都不属于这种类型. 比方说, 不能直接应用概率的频率解释某一对熟人在两年之内结婚的概率, 或一项特殊的医学科研项目在一

个特定时期内攻克某种疾病的概率.

1.2.2 概率的古典解释

概率的古典解释是基于等可能结果(equally likely outcomes)的概念. 比方说，投掷一枚硬币可能会出现两种可能的结果：正面或反面. 如果假定这两种结果出现的机会均等，它们必定有相同的概率. 由于概率的总和是1，所以出现正面和反面的概率必为1/2. 更一般地，如果某个过程的结果必是 n 种不同结果中的一种，并且这 n 种结果以相同的几率发生，那么每种结果的概率就是 $1/n$.

当试图从古典解释中发展概率的形式上的定义时，出现了两个根本性的难题. 首先，等可能结果的概念本质上是基于我们将要定义的概率的概念. 两个可能结果等可能出现的陈述和两个可能结果有相同概率的说法是一样的. 其次，没有一种系统的方法来设定等可能结果的概率. 当掷一枚硬币，或掷一枚均匀的骰子，或从一副充分洗过的牌中抽取一张的时候，由于过程的特性经常会使我们认为可能出现的不同结果的几率相等. 然而当问题是猜测一对熟人是否将会结婚或一个科研项目是否会成功时，认为可能出现的结果几率相等就不再是古典的了，这些结果的概率就需要一种不同的方法来设定.

3

1.2.3 概率的主观解释

根据概率的主观(或个人)解释，一个人对某个过程可能出现结果的设定取决于这个人对结果出现可能性的主观判断. 这个判断是基于每个人对这个过程的信念和信息而做出的. 另一个具有不同信念或掌握不同信息的人，对同一个结果也许会给出不同的概率. 基于这个原因，称某个人对一个结果的概率为主观概率比称其为真实概率(true probability)要更恰当.

作为该解释的一个说明，我们假设抛掷了一枚硬币. 一个对硬币本身或抛硬币方式一无所知的人认为出现正面或反面结果的可能性相等. 该人就会对得到正面的可能性赋以1/2的主观概率；然而那个实际抛硬币的人也许会感到正面更可能出现. 为了让普通人能设定结果出现的主观概率，他们会用很多形式来强调他们的观点. 比方说，假设他们认为抛硬币时正面出现的概率与从四张红牌和一张黑牌充分洗过后抽出一张红牌的概率相等，因为他们设定抽到红牌的概率为4/5，他们也应该设定抛一枚硬币时结果为正面的概率为4/5.

可以将概率的这种主观解释公式化. 一般来说，若人们对结果的不同组合的相对可能性的判断满足某种相容性要求，即表明可以唯一地确定不同可能事件的主观概率. 然而，这种主观解释遇到两个困难. 首先，要求人们对无数事件相对可能性的判断必须完全相容，且互相不矛盾，而这似乎不是人类所能做到的. 其次，主观解释没有提供两个或多个科学家一起工作，且对他们共同感兴趣的某个科研领域的知识状况而得到相同评价的客观根据.

另一方面，概率主观解释的认识对于强调科学的一些主观方面有益处. 一个科学家对于某个不确定结果的概率的评价最终一定是他自己基于所有已知资料的评价. 既然这位科学家可能把这种情况或过去类似的结果出现的相对概率都考虑在内，也许这个评价部分来自于对概率的频率理解. 也许这位科学家会把等可能出现的结果的总数考虑进去，那么这个评价也许就来自于对概率的古典解释，然而无论哪种情况，对数字化的概率的最终设定是科学家自己的责任.

4

在一个科学家从一系列可供选择的课题中选择研究课题时，在开展这项课题时所选择的实验里，在从实验数据中得出的结论中，都能显示出科学的主观特性. 概率和统计的数学理论在这些选择、决策和结论中起着重要的作用.

注： 概率理论不依赖于解释. 本书的 1 ~ 5 章介绍了概率的数学理论，而没有考虑围绕概率这个术语不同解释的争论. 如果将在特殊问题中用哪种概率的解释忽略不计，则该理论是正确的且能够充分利用. 本书中所提到的理论和技巧可以对实际实验设计和分析的几乎所有方面提供有价值的指导和工具.

1.3 试验和事件

概率将是我们定量地界定某件事可能发生的方法(依据 1.2 节关于概率解释的一种叙述). 本节中我们给出应用概率情形类别的一些例子.

1.3.1 试验类型

当进行一项试验时，概率理论适合于可能出现的不同结果，也适合于可能发生的事件. 术语“试验”在概率理论中是用来描述每一个不能预先确知结果的过程. 现在给出一些试验的例子如下：

1. 在将一个硬币抛 10 次的试验中，试验者可能想确定正面至少出现 4 次的概率.

2. 从一个装有相似零件的仓库中选出 1000 个晶体管样品，对每一个所选样品都要进行检测，试验者可能想确定所选晶体管中有不多于一个次品的概率.

3. 在连续 90 个中午对某固定地点测量气温的试验中，试验者可能想确定在这段时间中平均气温低于某个值的概率.

4. 从托马斯·杰弗逊的生平信息中，试验者可能想确定杰弗逊生于 1741 年的概率.

5. 在评价某一时间一项工业研究发展计划的过程中，试验者可能想确定在未来几个月内成功开发出新产品的概率.

5

依据这些术语的常用意义，从这些例子中可以看出，一项试验的可能结果也许是随机的，也许是确定的. 每项试验的有趣之处在于能在试验之前预测到它的每一个可能出现的结果，并可以确定所感兴趣的不同结果组合的概率.

1.3.2 概率的数学理论

正如1.2节所述，对很多试验结果的一些概率认定的合理意义和解释是有争议的. 然而，一旦确定了一项试验的一些简单结果的概率，所有权威们就会完全同意概率的数学理论为进一步研究这些概率提供合适的方法论. 在关于概率的几乎所有数学理论的著作中，从最基础的教科书到最高级的研究，都涉及下面两个问题：(i) 从一项试验的每一个可能出现结果的特定概率中确定某些事件概率的方法；(ii) 当得到额外的相关信息时修正概率事件的方法.

这些方法是基于标准数学技术的. 本书前五章的目的就是介绍这些技术，将其合在一起构成了概率的数学理论.

1.4 概率的定义

首先给出概率的数学定义，再给出容易从定义推导出的几个有用的结果.

我们称一个试验的所有可能结果组成的集合为该试验的样本空间(sample space)，通常记为 S. 每个结果可被认为是样本空间中的点(point)，或集合中的元素(element).

记空集合(即不含任何元素的集合)为 $\varnothing$；记两个集合 A 和 B 的并(即包含 A, B 中所有元素的新集合)为 $A \cup B$；记集合 A 和 B 的交(即既属于 A 又属于 B 的元素组成的新集合)为 $A \cap B$；而记集合 A 的补(即样本空间 S 中不属于 A 的元素组成的集合)为 A^c.

1.4.1 公理和基本定理

在本节我们将给出概率的数学定义或公理化定义. 在一个试验中，有必要给样本空间 S 中每个事件 A 设定数 $\Pr(A)$，该数表示事件 A 发生的概率. 为了满足概率的数学定义，设定的数 $\Pr(A)$ 必须满足三个特定的公理. 这些公理能确保我们直观上希望数 $\Pr(A)$ 具备1.2节所描写的概率性质. [12]

第一个公理阐明任何一个事件的概率必定非负.

公理1 对任意事件 A，$\Pr(A) \geqslant 0$.

第二个公理阐明若某事件肯定要发生，则该事件的概率为1.

公理2 $\Pr(S) = 1$.

在给出公理3之前我们首先讨论不相交事件的概率. 若两事件不相交，我们很自然地假定至少有一个事件发生的概率应是它们各自概率之和. 事实上，这种概率的可加性(additive property)也适合于任何有限个不相交事件的集合，甚至也适合于无限个不相交事件序列. 若假定可加性仅适合于有限个不相交事件，我们就不能肯定这个特性也适合于无限个不相交事件列. 然而，若假定任何一个无限不相交事件

列具有概率可加性，那么有限个不相交事件一定也具有这个性质(正如我们将要证明的那样). 综合以上考虑给出了第三个公理.

公理 3 对于任何无限不相交事件列 $A_1, A_2, \cdots$，有 $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$.

现在可以将概率的数学定义表述为：样本空间 S 上的概率分布，或简称为概率，是设定的满足公理 1、公理 2 和公理 3 的数 $\Pr(A)$.

下面根据公理 3 导出两个重要的推论. 首先证明若某事件不可能发生，则它的概率必定为 0.

定理 1.4.1 $\Pr(\varnothing) = 0$.

证明： 考虑无限事件列 $A_1, A_2, \cdots$，满足 $A_i = \varnothing, i = 1, 2, \cdots$，换言之，序列中每个事件都是空集$\varnothing$. 因为$\varnothing \cap \varnothing = \varnothing$，因此是无限不相交事件列. 更进一步，$\bigcup_{i=1}^{\infty} A_i = \varnothing$. 因

13 此，由公理 3 得：$\Pr(\varnothing) = \Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i) = \sum_{i=1}^{\infty} \Pr(\varnothing)$. 该方程表明数 $\Pr(\varnothing)$被无限次相加后其和仍是 $\Pr(\varnothing)$，满足该性质的实数只有零. ■

现在可以证明公理 3 中对无限不相交事件列的可加性，对任何有限不相交事件也成立.

定理 1.4.2 对任意 n 个不相交事件 $A_1, \cdots, A_n$ 的有限序列有

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i).$$

证明： 考虑事件 $A_1, A_2, \cdots$的无限序列，其中 $A_1, \cdots, A_n$ 就是 n 个给定的不相交事件，当 $i > n$ 时，$A_i = \varnothing$. 则无限列中的事件互不相交且有 $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$. 从而根据公理 3 有：

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i) = \sum_{i=1}^{n} \Pr(A_i) + \sum_{i=n+1}^{\infty} \Pr(A_i)$$

$$= \sum_{i=1}^{n} \Pr(A_i) + 0 = \sum_{i=1}^{n} \Pr(A_i).$$ ■

1.4.2 概率的进一步性质

根据上面的公理和定理，现推导概率分布的另外四个一般性质. 由这些性质的基本特性，我们就以定理的形式给出，它们都很容易证明.

定理 1.4.3 对任意事件 A，有 $\Pr(A^c) = 1 - \Pr(A)$.

证明： A 和 A^c 是不相交的事件且有 $A \cup A^c = S$，由定理 1.4.2 得 $\Pr(S) = \Pr(A) +$

14 $\Pr(A^c)$. 根据公理 2 有 $\Pr(S) = 1$，则 $\Pr(A^c) = 1 - \Pr(A)$. ■

定理 1.4.4 如果 $A \subset B$，则 $\Pr(A) \leqslant \Pr(B)$.

证明： 如图 1-1 所示，可将事件 B 理解为两个不相交事件 A 和 BA^c 的并. 从而有 $\Pr(B) = \Pr(A) + \Pr(BA^c)$. 又 $\Pr(BA^c) \geqslant 0$，则 $\Pr(B) \geqslant \Pr(A)$. ■

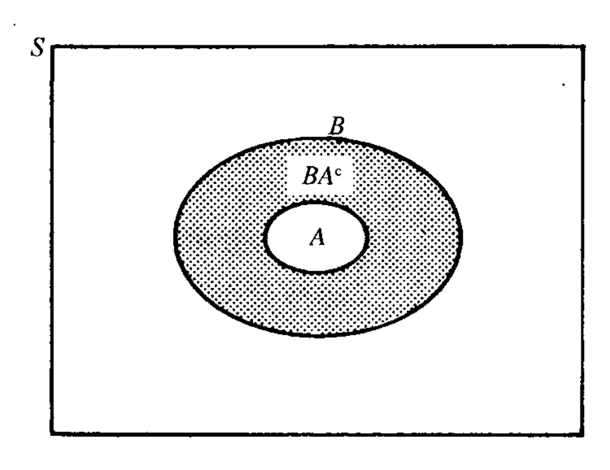


图 1-1 $B = A \cup (BA^c)$

定理 1.4.5 对任意事件 A，有 $0 \leqslant \Pr(A) \leqslant 1$.

证明： 由公理 1 可知 $\Pr(A) \geqslant 0$. 因为对每个事件 A，有 $A \subset S$，那么根据公理 2 和定理 1.4.4 有 $\Pr(A) \leqslant \Pr(S) = 1$. ■

定理 1.4.6 对任意两个事件 A 和 B，有

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB). \tag{1.4.1}$$

证明： 如图 1-2 所示，$A \cup B = (AB^c) \cup (AB) \cup (A^cB)$. 因为该等式右边的三个事件不相交，根据定理 1.4.2 有

$$\Pr(A \cup B) = \Pr(AB^c) + \Pr(AB) + \Pr(A^cB). \tag{1.4.2}$$

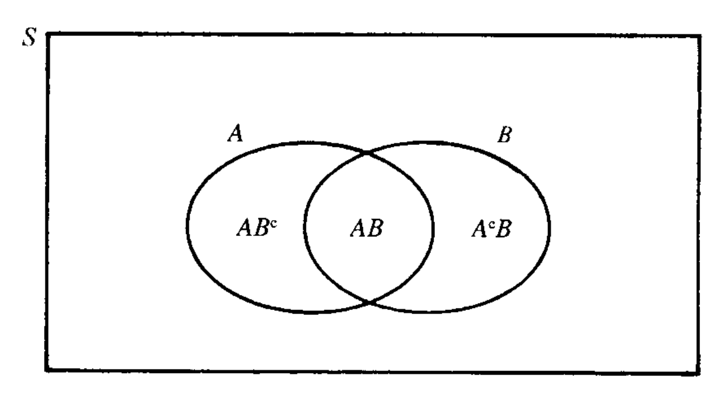


图 1-2 $A \cup B$ 的分解

15

而且由图 1-2 还可以看出

$$\Pr(A) = \Pr(AB^c) + \Pr(AB)$$

和

$$\Pr(B) = \Pr(AB) + \Pr(A^cB).$$

利用最后这两个等式计算(1.4.1)式右边的值：

$$\begin{aligned}\Pr(A) + \Pr(B) - \Pr(AB) &= \Pr(AB^c) + \Pr(AB) + \Pr(AB) + \Pr(A^cB) - \Pr(AB) \\ &= \Pr(AB^c) + \Pr(AB) + \Pr(A^cB),\end{aligned}$$

即为(1.4.2)式右边. ■

例 1.4.1　疾病诊断　一个患有咽喉炎和低烧的病人去诊所看病. 检查后，医生发现病人或是细菌性感染或是病毒性感染或两者兼而有之. 医生认为病人是细菌性感染的概率为 0.7，而是病毒性感染的概率为 0.4. 那么病人同时感染两者的概率是多少?

设 B 表示病人是细菌性感染的事件; V 表示病人是病毒性感染的事件. 于是 $\Pr(B)=0.7$，$\Pr(V)=0.4$，且有 $S=B\cup V$. 我们的目的是求出 $\Pr(BV)$. 利用定理 1.4.6，即

$$\Pr(B\cup V)=\Pr(B)+\Pr(V)-\Pr(BV), \tag{1.4.3}$$

因为 $S=B\cup V$，则(1.4.3)的左边是 1，右边前两项分别是 0.7 和 0.4. 结果为 $1=0.7+0.4-\Pr(BV)$，从而可得到 $\Pr(BV)=0.1$，即为病人两种感染都有的概率. ◀

例 1.4.2　水电需求量　一个承包商承建一栋办公楼，需要计划水电的需求（水泵、水管和电线的用量），经过向潜在的客户咨询并考查了历史数据，承包商确认电的需求约在每天 1 百万到 1.5 亿千瓦时，水的需求约在每天 4 ~ 200 千加仑之间. 考虑水电需求量的所有可能的组合，其形状如图 1-3 所示，图 1-3 中的阴影部分表示了办公楼水电需求量试验的样本空间，这个样本空间包括了无穷多个点，事实上样本空间是不可数的.

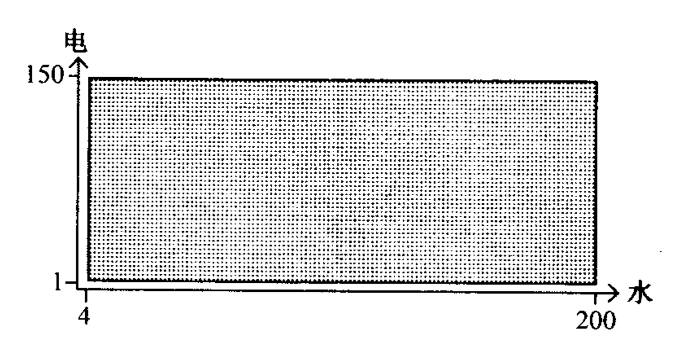


图 1-3　水电需求量的样本空间

在这个样本空间上有很多办法定义概率，一种比较简单的办法是设事件 E 的概率和 E 的面积成比例. S(样本空间)的面积是 $(150-1)\times(200-4)=29\,204$，于是 $\Pr(E)$ 就等于 E 的面积除以 29 204 所得的商. 例如，假设承包商感兴趣于高需求量. 让 A 表示水的需求量至少是 100 的集合，而让 B 表示电的需求量至少是 115 的集合，并假设这些值即为所谓的高需求量，在图 1-4 中用不同的阴影表示事件 A 和 B. A 的面积是 $(150-1)\times(200-100)=14\,900$，$B$ 的面积是 $(150-115)\times(200-4)=6860$，于是有，$\Pr(A)=\dfrac{14\,900}{29\,204}=0.5102$，　$\Pr(B)=\dfrac{6860}{29\,204}=0.2349$. 记两个事件相交部分为 AB，它对应的面积是 $(150-115)\times(200-100)=3500$，于是 $\Pr(AB)=3500/29\,204=0.1198$. 如果承包商希望计算两个需求中至少有一个是高需求的概率，则根据定理 1.4.6 可得相应的概率为

$$\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(AB)=0.5102+0.2349-0.1198=0.6253.$$

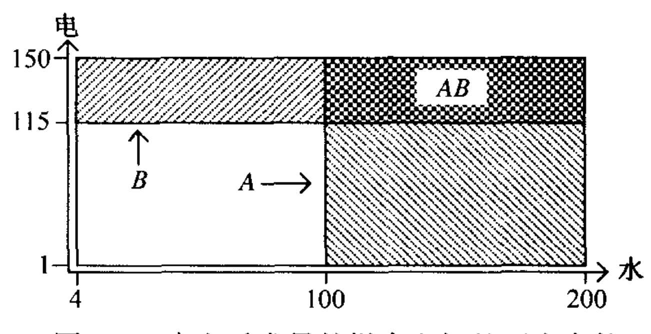


图 1-4 水电需求量的样本空间的两个事件 ◀

注： 零概率并不意味着不可能. 当一个事件概率为零时并不意味着该事件不可能发生. 在例 1.4.2 中，有许许多多概率为零的事件，但它们不都是不可能发生事件. 例如，对任何 x，水的需求量为 x 的事件对应于图 1-3 中一条线段. 既然线段的面积是 0，则每个这样的事件的概率为 0，但这些事件不都是不可能的. 事实上，如果诸如{水的需求量为 x}这类事件不可能的话，那么水的需求量就不能取任何一个值. 设 $\varepsilon>0$,则事件{水的需求量介于 $x-\varepsilon$ 和 $x+\varepsilon$ 之间}有正的概率，但当 $\varepsilon\to 0$ 时却趋于 0.

1.4.3 小结

我们已通过三个公理给出了概率的数学定义. 这些公理要求每个事件的概率非负；样本空间的概率为 1；无限不相交事件列的概率等于各事件概率的总和. 要求记住的要点有：

- 如果事件 $A_1,\cdots,A_k$ 是不相交的，则有 $\Pr\left(\bigcup_{i=1}^{k} A_i\right)=\sum_{i=1}^{k}\Pr(A_i)$.
- $\Pr(A^c)=1-\Pr(A)$.
- 如果 $A\subset B$，则 $\Pr(A)\leqslant\Pr(B)$.
- $\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(AB)$.

17

概率是怎么确定的无关紧要，只要它满足这三条公理，则必定满足上述关系及后面证明的结论.

习题

1. 从一个装有红色、白色、蓝色、黄色和绿色球的箱子中挑选一个球，如果挑选出的球是红色的概率为 1/5，是白色的概率为 2/5，那么是蓝色、黄色或绿色的概率为多少?
2. 从班里选出一名学生. 如果是男生的概率为 0.3，那么是女生的概率是多少?
3. 设两个事件 A 和 B 的概率分别是 1/3 和 1/2. 试求在以下条件下 $\Pr(BA^c)$的值：(a) A 和 B 是不相交的；(b) $A\subset B$；(c) $\Pr(AB)=1/8$.
4. 设在统计课考试中学生 A 不及格的概率是 0.5，学生 B 不及格的概率是 0.2，两人同时不及格的概率是 0.1，那么至少有一个不及格的概率是多少?

5. 在习题4的条件下，两人都考及格的概率是多少？
6. 在习题4的条件下，两人中只有一个人不及格的概率是多少？
7. 考虑两个事件 A 和 B 满足 $\Pr(A)=0.4$，$\Pr(B)=0.7$，试求 $\Pr(AB)$ 的最大值和最小值并写出达到这些值的条件.
8. 在某一个城市里，有50%的家庭订阅晨报，有65%的家庭订阅晚报，而有85%的家庭至少订阅其中的一份. 问同时订阅两种报纸的家庭占的百分比是多少？
9. 试证明在任意两个事件 A 和 B 中只有一件发生的概率是 $\Pr(A)+\Pr(B)-2\Pr(AB)$.
10. 对任意两个事件 A 和 B，试证明：$\Pr(A)=\Pr(AB)+\Pr(AB^c)$.
11. 现从包含 $0\leqslant x\leqslant 1$，$0\leqslant y\leqslant 1$ 的所有点的正方形 S 中任选一点 (x,y). 设所选的点属于 S 的某个子集的概率等于该子集的面积. 求下面每个子集的概率：(a) 满足 $\left(x-\frac{1}{2}\right)^2+\left(y-\frac{1}{2}\right)^2\geqslant\frac{1}{4}$ 的点组成的子集；(b) 满足 $\frac{1}{2}<x+y<\frac{3}{2}$ 的点组成的子集；(c) 满足 $y\leqslant 1-x^2$ 的点组成的子集；(d) 满足 $x=y$ 的点组成的子集.
12. 设 $A_1,A_2,\cdots$，是一个任意的无限事件列，再设 $B_1,B_2,\cdots$，是另一个如下定义的无限事件列 $B_1=A_1$，$B_2=A_1^cA_2$，$B_3=A_1^cA_2^cA_3$，$B_4=A_1^cA_2^cA_3^cA_4$，$\cdots$，试证明：$\Pr\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}\Pr(B_i)$，对 $n=1,2,\cdots$，和 $\Pr\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\Pr(B_i)$ 成立.
13. 对所有事件 $A_1,\cdots,A_n$，求证：$\Pr\left(\bigcup_{i=1}^{n}A_i\right)\leqslant\sum_{i=1}^{n}\Pr(A_i)$.
14. 人类血型的一个简化模型包括4种血型：A、B、AB、O型和2种抗体：抗A和抗B. 抗体依据血型的不同而与人的血液以不同的形式发生作用，抗A只与A和AB型血发生作用，不与B和O型血作用，抗B只与B和AB型血发生作用，不与A和O型血作用. 假设一个人的血型是O型的概率为0.5，是A型的概率为0.34，是B型的概率为0.12.
   a. 求每个抗体与此人的血发生作用的概率；
   b. 求此人的血与两种抗体都起作用的概率.

18

1.5　有限样本空间

在研究概率的试验中，最简单的是只包含有限多个可能结果的试验. 本节将用几个例子说明1.4节中讲到的有限样本空间的重要概念.

1.5.1　概率的要求条件

本节将考虑只含有有限个可能结果的试验. 也就是说，考虑这些试验的样本空间 S 仅包含有限个样本点 $s_1,s_2,\cdots,s_n$. 在该种类型的试验里，样本空间 S 上的概率分布的特征表现为每一个样本点 $s_i\in S$ 都对应一个概率 p_i. p_i 就是试验结果 $s_i(i=1,\cdots,n)$ 发生的概率. 为了满足概率分布的公理，$p_1,\cdots,p_n$ 必须满足下面两个条件：

$$p_i \geqslant 0,\ i=1,\ \cdots,\ n \text{ 和 } \sum_{i=1}^{n} p_i = 1.$$

事件 A 发生的概率就等于 A 中所有结果 s_i 对应的概率 p_i 之和.

例 1.5.1 线断裂 考虑以下试验，通过测试过程来检验五条长度不等的线中哪一根最先断. 假设这五根线的长度分别是 1、2、3、4 和 5 英寸，每一根最先断的可能性与它的长度成正比. 我们将确定最先断的线的长度不超过 3 英寸这个事件的概率.

在这个例子中，令 s_i 表示长度为 $i(i=1,\cdots,5)$ 的线先断这个结果，则 $S=\{s_1,\cdots,s_5\}$，$p_i=\alpha i(i=1,\cdots,5)$ 这里 α 是比例因子. 显然一定有 $p_1+\cdots+p_5=1$，并且有 $p_1+\cdots+p_5=15\alpha$，因此 $\alpha=1/15$. 若 A 表示最先断的线的长度不超过 3 英寸这个事件，则 $A=\{s_1,\ s_2,\ s_3\}$，$\Pr(A)=p_1+p_2+p_3=\frac{1}{15}+\frac{2}{15}+\frac{3}{15}=\frac{2}{5}$. ◀

19

1.5.2 简单样本空间

一个包含 n 个结果 $s_1,s_2,\cdots,s_n$ 的样本空间 S，如果对应于每个结果 $s_1,s_2,\cdots,s_n$ 的概率都是 $1/n$，则称它是一个简单样本空间. 若这个简单样本空间里的事件 A 包含 m 个结果，则 $\Pr(A)=\frac{m}{n}$.

例 1.5.2 掷硬币 假设同时掷三个相同的硬币. 确定正好得到两个正面的概率.

为了描述样本空间方便起见，不论试验者能否区分这三个硬币，都假设可以将它们区分出来. 我们称三次抛掷结果分别为第一个硬币的结果、第二个硬币的结果和第三个硬币的结果. 如果用 H 表示正面，T 表示反面，则样本空间包含如下 8 个可能的结果：

s_1：HHH，　s_2：THH，　s_3：HTH，　s_4：HHT，

s_5：HTT，　s_6：THT，　s_7：TTH，　s_8：TTT.

基于硬币是均匀的假设，有理由认为这个样本空间是简单的，并且这 8 个结果中的每一个发生的概率都是 1/8. 由上述表中可以看出，其中有三个结果中正好是两个正面. 因此，得到两个正面的概率是 3/8. ◀

值得注意的是，若考虑没有正面、一个正面、两个正面、三个全是正面这些结果的可能性时，则假设样本空间仅包含这四种结果就是合理的. 但由于这四种结果并不是等可能的，这个样本空间就不是简单的.

例 1.5.3 遗传 人类的遗传特性是由染色体特定位置上的物质所决定的，每个人都从父母那里得到 23 对染色体，并且这些染色体是自然成对的，每对均分别来自父亲和母亲. 此时可将基因看作每个染色体对中的一部分，单独一个基因或者一个基因组决定着比如血型、发色这样的遗传特性. 称组成染色体对的两个位置上的物质为等位基因，称等位基因的每个不同的组合为基因型.

下面考虑只有两个不同等位基因 A 和 a 的情况. 假设父母双方都有基因型 Aa, 也即父母各方的一对染色体中, 其中一个有 A, 另一个有 a. (这里我们把组成等位基因相同但排列顺序不同的基因型看作是相同的基因型. 例如 aA 和 Aa 是同一个基因型. 但在计算概率的中间步骤中,如例 1.5.2 区别三个硬币的方式来区分两个染色体是很方便的.)该对父母的后代们的基因型可能是什么? 在所有的可能结果中, 如果父母双方贡献等位基因是等可能性的, 那么不同基因型出现的概率是多少呢?

首先, 既然已经假设双方等可能性地贡献等位基因, 就应该能区分后代从各方获得的是哪一个等位基因. 然后, 把生成同一种基因型的结果结合起来, 来自父母的可能组合就是:

	母亲	
父亲	A	a
A	AA	Aa
a	aA	aa

20

因此, 后代可能会出现三种基因型 AA, Aa 和 aa. 既然已假定每种组合是等可能的, 则上面表格里的四个单元的概率都是 1/4. 由于表中有两个结果是 Aa, 该基因型的概率就是 1/2 而其余两个的概率就是 1/4. ◀

例 1.5.4　掷双骰子　考虑掷两个质地均匀骰子的试验. 计算每次出现的两个点数之和的每个可能值的概率.

尽管试验者为了观察点数之和的值而无须区别这两个骰子, 但假设这两个骰子是不同的可方便地描述出该实验的简单样本空间. 基于这个假设, 用一个数对 (x,y) 来表示样本空间 S 中的每个结果, 其中 x 和 y 分别是第一个和第二个骰子掷出来的点数. 因此, S 就包含下面 36 个结果:

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6)\\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6)\\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6)\\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6)\\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6)\\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

显然能得到 S 是一个简单样本空间且每个结果的概率都是 1/36.

定义 P_i 是当和为 i 时的概率, $i=2,3,\cdots,12$. S 中和等于 2 的结果只有 $(1,1)$. 因此, $P_2=1/36$. 和等于 3 的结果可能是 $(1,2)$ 或 $(2,1)$, 因此 $P_3=2/36=1/18$. 按这种方式算下去, 就得到所有可能结果的概率如下:

$$P_2=P_{12}=\frac{1}{36},\quad P_5=P_9=\frac{4}{36},\quad P_3=P_{11}=\frac{2}{36},$$

$$P_6=P_8=\frac{5}{36},\quad P_4=P_{10}=\frac{3}{36},\quad P_7=\frac{6}{36}.$$ ◀

1.5.3 小结

简单样本空间 S 是其中每个结果发生的概率都相等的有限样本空间. 如果简单样本空间 S 中有 n 个结果，则每个结果的概率都是 $1/n$. 在一个简单样本空间中事件 E 的概率即为 E 中的结果数除以 n. 在以后的三节里，将会介绍几种有用的方法来计算不同事件中出现的结果的数目.

21

习题

1. 掷 2 个质地均匀的骰子，求 2 个数字的和是奇数的事件的概率.
2. 掷 2 个质地均匀的骰子，求 2 个数字的和是偶数的事件的概率.
3. 掷 2 个质地均匀的骰子，求 2 个数字的差小于 3 的事件的概率.
4. 一个学校有 1、2、3、4、5 和 6 年级. 2、3、4、5 和 6 年级有相同数量的学生，但 1 年级是其他任一年级学生数的 2 倍. 如果从该校学生中随机抽出一名，则这名学生刚好在 3 年级的概率是多少?
5. 在第 4 题的条件下，抽出的那名学生所在年级是奇数的概率是多少?
6. 掷 3 个均匀的硬币，3 个硬币正反恰好一致的概率是多少?
7. 在例 1.5.3 的条件下，现假设两对父母的基因型分别为 Aa 和 aa，求后代可能的基因型和每个基因型的概率，仍假设父母双方贡献等位基因的可能结果的概率是相等的.
8. 进行以下试验：掷一个均匀硬币然后掷一个均匀的骰子.
   a. 描述该试验的样本空间;
   b. 硬币是正面且骰子点数是奇数的概率是多少?

22

1.6 多项式系数

记 n 个不同元素的排列方法有 $n!$ 种，记从 n 个不同元素中选取 k 个的选法有 $\binom{n}{k}$ 种，由组合数学可知，

$$n! = n\times(n-1)\times(n-2)\times\cdots\times2\times1 \qquad \binom{n}{k}=\frac{n!}{k!\,(n-k)!}$$

$\binom{n}{k}$ 又称为二项式系数，因为它出现在二项式定理中.

我们学习过如何计算把一个有限集合分解成多于两个不相交子集的方法的个数. 当分割的结果是由从不同类型中选出的一些成分所组成时，这种推广是很有用的.

现将 n 个不同的元素分成 k 个不同的组. 设第 j 组含有 n_j 个元素，$j=1,2,\cdots,k$，且有 $n_1+n_2+\cdots+n_k=n$. 现在想知道把 n 个元素分成 k 个组一共有多少种不同的分

法. 从 n 个元素中选出第一组的 n_1 个元素有 $\binom{n}{n_1}$ 种不同的方法；接着，从剩下的 $n-n_1$ 个元素中选出第二组的 n_2 个元素有 $\binom{n-n_1}{n_2}$ 种不同的方法，因此选出前两组的元素共有 $\binom{n}{n_1}\binom{n-n_1}{n_2}$ 种不同的方法. 而从剩下的 $n-n_1-n_2$ 个元素中选出第三组的 n_3 个元素有 $\binom{n-n_1-n_2}{n_3}$ 种方法. 这样一来选出前三组共有 $\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}$ 种不同的方法.

这样继续下去，当选出前 $k-2$ 组后还剩下 $n_{k-1}+n_k$ 个元素，从中选出 n_{k-1} 个元素共有 $\binom{n_{k-1}+n_k}{n_{k-1}}$ 种不同的方法. 而剩下的 n_k 个元素自然组成最后一组即第 k 组. 因此把 n 个元素分成 k 组共有 $\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_{k-1}+n_k}{n_{k-1}}$ 种不同的方法. 把二项式系数写成阶乘的形式，则该乘积变为 $\frac{n!}{n_1!\ n_2!\ \cdots n_k!}$ 这样简单的形式，我们把它记作 $\binom{n}{n_1,n_2,\cdots,n_k}$，并称为多项式系数，之所以这样叫是因为它出现在多项式定理里面，这个定理的证明将留作习题.

35

定理 1.6.1　多项式定理　对数 $x_1,\cdots,x_k$ 和每个正整数 n，

$$(x_1+\cdots+x_k)^n=\sum\binom{n}{n_1,n_2,\cdots,n_k}x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k},$$

在这个和式中，包含所有可能的满足 $n_1+n_2+\cdots+n_k=n$ 的非负整数 $n_1,\cdots,n_k$ 的组合. ■

多项式系数是二项式系数的推广. 当 $k=2$ 时，多项式定理就是二项式定理，而多项式系数就变成二项式系数. 特别地有，

$$\binom{n}{k,n-k}=\binom{n}{k}.$$

例 1.6.1　选择委员会　把一个组织中的 20 个成员分配到 A，B 和 C 三个委员会中，使委员会 A 和 B 各有 8 名成员，C 有 4 名成员，求不同的成员分配方案有多少种.

这里要求的是多项式系数，其中 $n=20$，$k=3$，$n_1=n_2=8$，$n_3=4$. 因此解是

$$\binom{20}{8,8,4}=\frac{20!}{(8!)^2 4!}=62\ 355\ 150.$$ ◀

多于两种不同类型的元素的排列

与我们用二项式系数来表示只有两种不同类型元素的排列数目一样，我们用多项式系数来表示集合中有 $k(k\geq 2)$ 种不同类型的元素排列的数目. 例如，把 n 个有 k 种不同颜色的单色球排成一行，得到 n_j 个 j 颜色的球 $(j=1,\cdots,k)$，满足 $n_1+n_2+\cdots+n_k=n$. 因此 n 个球的每种排列方法就对应于一行中 n 个空位分别由第 1 种颜色的球占据 n_1 个位置，第 2 种颜色的球占据 n_2 个位置等等. 因此，这 n 个球的各种不同方法的排列方式的总数一定是

$$\binom{n}{n_1,n_2,\cdots,n_k}=\frac{n!}{n_1!\ n_2!\ \cdots n_k!}.$$

例 1.6.2 掷骰子 掷 12 个骰子，求结果中 6 种点数每种恰好出现两次的概率.

把样本空间的每次结果看作有 12 个数字的序列，其中该序列的第 i 个数就表示第 i 次掷的结果. 因此，在 S 中会有 6^{12} 种可能的结果，且认为出现每种结果都是机会均等的. 结果中包含数字 $1,2,\cdots,6$ 中的每一个恰好出现两次的数目，就等于这 12 个数字的不同组合方式的数目. 通过估计满足 $n=12,k=6,n_1=n_2=\cdots=n_6=2$ 的多项式系数来决定这个数. 因此，所求的就是

[36]

$$\binom{12}{2,2,2,2,2,2}=\frac{12!}{(2!)^6},$$

且所求的概率 p 是

$$p=\frac{12!}{2^6 6^{12}}=0.0034.$$

◀

例 1.6.3 玩牌 一副 52 张的牌有 13 张红桃. 洗牌后，分发给四个玩家 A，B，C 和 D，使每人收到 13 张牌，求玩家 A 有 6 张红桃，玩家 B 有 4 张红桃，玩家 C 有 2 张红桃，玩家 D 有 1 张红桃的概率 p.

把 52 张牌平均分给 4 个玩家，使每人手中有 13 张牌，这样不同的分法总共有 N 种

$$N=\binom{52}{13,13,13,13}=\frac{52!}{(13!)^4}.$$

可设这样的每种分法都是等可能的，现在我们要计算的是使每位玩家得到所要求数量红桃的不同分法数目 M，把红桃分给 A、B、C 和 D 四个玩家，使得他们分别收到红桃的数目是 6、4、2 和 1 的不同分法是

$$\binom{13}{6,4,2,1}=\frac{13!}{6!\ 4!\ 2!\ 1!}.$$

并且把其他 39 张牌分给四个玩家，使他们每人手中总共有 13 张牌，这样不同的分法有

$$\binom{39}{7,9,11,12}=\frac{39!}{7!\ 9!\ 11!\ 12!}.$$

因此，

$$M=\frac{13!}{6!\ 4!\ 2!\ 1!}\cdot\frac{39!}{7!\ 9!\ 11!\ 12!}.$$

所求的概率 p 是

$$p=\frac{M}{N}=\frac{13!\ 39!\ (13!)^4}{6!\ 4!\ 2!\ 1!\ 7!\ 9!\ 11!\ 12!\ 52!}=0.001\ 96.$$

本问题还有另一种解法，这副牌中红桃所占 13 个位置的不同组合数是 $\binom{52}{13}$.

[37] 若玩家 A 收到 6 张红桃，那么这 6 张红桃占 A 手中 13 张牌的位置会有 $\binom{13}{6}$ 种可能. 同样的，若玩家 B 得到 4 张红桃，那么在 B 手里的 13 张牌中，红桃就会有 $\binom{13}{4}$ 种可能的位置组合. 玩家 C 就会有 $\binom{13}{2}$ 种可能的组合，玩家 D 会有 $\binom{13}{1}$ 种可能的组合. 因此，

$$p=\frac{\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}},$$

得出的结果和前一种方法是一致的. ◀

小结

多项式系数推广了二项式系数. 系数 $\binom{n}{n_1,\cdots,n_k}$ 就是把有 n 个元素的集合分成大小为 $n_1,\cdots,n_k$ 的不同子集的方法数，其中 $n_1+n_2+\cdots+n_k=n$. 它也是把有 k 种不同类型的 n 个元素排列成有 n_i 个是第 $i(i=1,\cdots,k)$ 种类型的排列方法数. 例 1.6.3 也反映了另外一个关于概率计算的重要信息：计算同样的概率可能有不止一种正确的方法.

习题

1. 三个调研员到一个有 21 户的社区进行民意调查，每个调研员去其中的 7 户. 问这样会有多少种不同的调查方式?
2. 把 18 颗红珠子、12 颗黄珠子、8 颗蓝珠子和 12 颗黑珠子串成一串. 问能串成多少种不同排列的颜色串?
3. 从有 300 个成员的组织中选取成员成立两个委员会. 若一个委员会需 5 名会员，另一个需 8 名会员，求有多少种不同的选取方式来组成这两个委员会?

4. 随机地排列字母 s、s、s、t、t、t、i、i、a、c，求拼成单词“statistics”的概率是多少？
5. 掷 n 个均匀的骰子，在满足 $n_1+n_2+\cdots+n_6=n$ 的条件下，求数字 j 恰好出现 n_j 次的概率是多少？
6. 掷 7 个均匀的骰子，则 6 个数字中每个出现至少一次的概率是多少？
7. 一副 25 张的牌，其中有 12 张红牌. 随机地把这 25 张牌分发给玩家 A、B 和 C，使 A 有 10 张、B 有 8 张、C 有 7 张牌. 求 A 有 6 张红牌、B 有 2 张红牌、C 有 4 张红牌的概率是多少？
8. 一副 52 张的牌，其中有 12 张“彩”牌. 随机地把这 52 张牌分发给四个玩家，使每个玩家手中有 13 张，求每位玩家手中有 3 张“彩”牌的概率是多少？
9. 一副 52 张的牌，其中有 13 张红牌，13 张黄牌，13 张蓝牌和 13 张绿牌. 随机地把这 52 张牌分给四个玩家，使每个玩家手中有 13 张，求每个玩家手中有 13 张同色牌的概率是多少？
10. 假设有 2 个叫 Davis 的男孩，3 个叫 Jones 的男孩，4 个叫 Smith 的男孩随意地坐在一排 9 座的座位上. 求在这一排座位上叫 Davis 的男孩刚好坐在前两个座位，叫 Jones 的男孩坐在接着的 3 个座位，叫 Smith 的男孩坐在最后 4 个座位的概率是多少？
11. 证明多项式定理.（提示：用归纳法证明）.

38

1.7 事件和的概率

概率的公理说明了如何去求不相关事件的和的概率. 定理 1.4.6 说明了如何去求任意两事件和的概率. 此定理可推广到任意有限个事件的和.

1.7.1 三个事件的和

现在再来考虑任意一个样本空间 S，它可能包含有限多个也可能包含无限多个结果，下面将研究可能由 S 中事件所确定的各种不同概率的更一般的一些性质. 在本节，特别要研究的是 n 个事件 $A_1,\cdots,A_n$ 的和 $\bigcup_{i=1}^{n} A_i$ 的概率.

如果事件 $A_1,\cdots,A_n$ 是不相交的，则知

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n}\Pr(A_i).$$

而且，对于任意的两个事件 A_1 和 A_2，不管是否相交，由 1.4 节的定理 1.4.6 可知

$$\Pr(A_1\cup A_2)=\Pr(A_1)+\Pr(A_2)-\Pr(A_1A_2).$$

现在来推广这个结果，先对三个事件再到任意有限个事件.

定理 1.7.1 对任意三个事件 A_1，A_2 和 A_3，有

$$\begin{aligned}\Pr(A_1\cup A_2\cup A_3)&=\Pr(A_1)+\Pr(A_2)+\Pr(A_3)\\&\quad-[\Pr(A_1A_2)+\Pr(A_2A_3)+\Pr(A_1A_3)]+\Pr(A_1A_2A_3).\end{aligned}\tag{1.7.1}$$

从这个方程可以看出，$\Pr(A_1\cup A_2\cup A_3)$的值可以通过求三个事件的概率和，减去任意两个事件的交的概率和，再加上三个事件交的概率.

证明：由图1-5知事件和$A_1 \cup A_2 \cup A_3$可以表示成7个不相交事件的和，分别用$p_1, \cdots, p_7$来表示如图1-5所示的7个不相交事件的概率，见图1-6. 则$\Pr(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{7} p_i$. 我们所要证明的就是等式(1.7.1)的右端也等于$\sum_{i=1}^{7} p_i$.

39

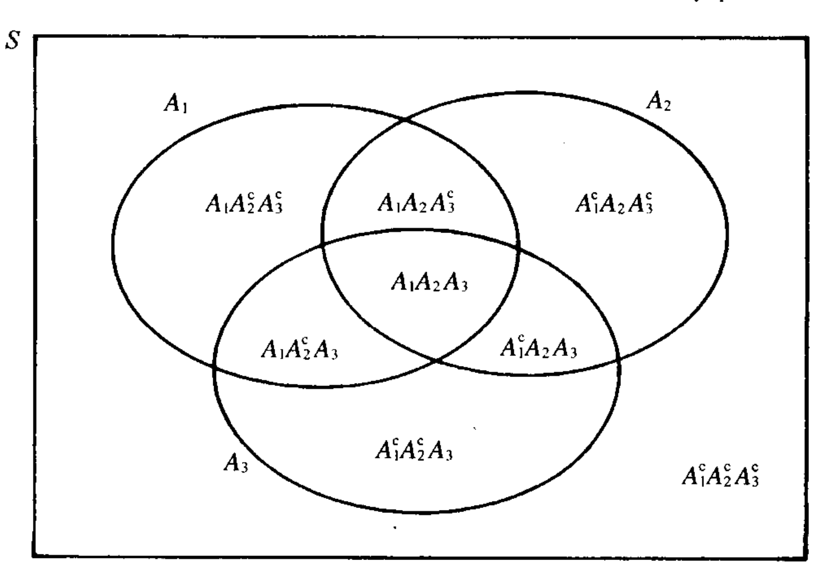


图1-5 由三个事件A_1，A_2和A_3决定的S的分割

从图1-6可以看出下列三式是正确的：

$$\begin{aligned}\Pr(A_1)+\Pr(A_2)+\Pr(A_3) &= (p_1+p_4+p_6+p_7)+(p_2+p_4+p_5+p_7) \\ &\quad +(p_3+p_5+p_6+p_7) \\ &= p_1+p_2+p_3+2p_4+2p_5+2p_6+3p_7, \\ \Pr(A_1A_2)+\Pr(A_2A_3)+\Pr(A_1A_3) &= (p_4+p_7)+(p_5+p_7)+(p_6+p_7) \\ &= p_4+p_5+p_6+3p_7, \\ \Pr(A_1A_2A_3) &= p_7.\end{aligned}$$

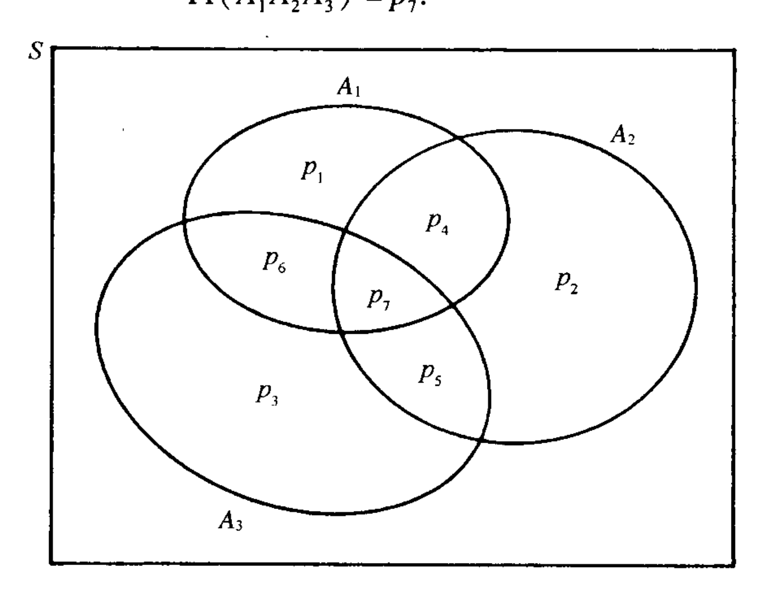


图1-6 三个事件并$A_1 \cup A_2 \cup A_3$的7个事件的概率

从这些关系即可得到等式(1.7.1)的右端等于 $\sum_{i=1}^{7} p_i$. ■

例 1.7.1 学生选课 有200名学生，其中137名学生报名数学班，50名学生报名历史班，124名学生报名音乐班. 除此之外，同时报名数学和历史班的学生是33名；同时报名历史和音乐班的学生是29名；同时报名数学和音乐班的是92名. 最后，同时报名三个班的学生是18名. 从这200名学生中随意挑取一名，求他至少报名这三个班中一个的概率.

40

记 A_1 表示挑到的学生报名数学班的事件；记 A_2 表示他报名历史班的事件；记 A_3 表示他报名音乐班的事件. 要解本题，首先要求 $\Pr(A_1 \cup A_2 \cup A_3)$ 的值. 从以上给出的数知，

$$\Pr(A_1) = \frac{137}{200}, \quad \Pr(A_2) = \frac{50}{200}, \quad \Pr(A_3) = \frac{124}{200},$$

$$\Pr(A_1A_2) = \frac{33}{200}, \quad \Pr(A_2A_3) = \frac{29}{200}, \quad \Pr(A_1A_3) = \frac{92}{200},$$

$$\Pr(A_1A_2A_3) = \frac{18}{200}.$$

从等式(1.7.1)可得 $\Pr(A_1 \cup A_2 \cup A_3) = 175/200 = 7/8$. ◀

1.7.2 有限个事件的和

对于任意有限个事件，下面的定理给出一个和等式(1.7.1)类似的方程：

定理 1.7.2 对任意 n 个事件 $A_1, \cdots, A_n$,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i<j} \Pr(A_iA_j) + \sum_{i<j<k} \Pr(A_iA_jA_k) - \sum_{i<j<k<l} \Pr(A_iA_jA_kA_l) + \cdots + (-1)^{n+1} \Pr(A_1A_2\cdots A_n). \tag{1.7.2}$$

证明：等式(1.7.2)的证明和等式(1.7.1)的证明类似，但此处，不能依据韦恩图. 假设把 $\bigcup_{i=1}^{n} A_i$ 分割成不相交的子集，使每个子集所包含的结果是 $A_1, \cdots, A_n$ 中某些特定的事件所共有的，并且不属于任何其他的事件. 则 $\Pr(\bigcup_{i=1}^{n} A_i)$ 就是这些不相交子集的概率的和. 要证定理1.7.2，必须证明在等式(1.7.2)的右端每个子集的概率只被计算了一次.

对于 k 的一个给定的值 $(1 \leq k \leq n)$，首先考虑结果中的子集 B 属于 $A_1, \cdots, A_k$ 中的每一个事件但不属于 $A_{k+1}, \cdots, A_n$ 中的任何事件. 即考虑

$$B = A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c.$$

因为 B 是 n 个事件 $A_1, \cdots, A_n$ 中正好 k 个事件的子集，那么在等式(1.7.2)的右端第一个和项中 $\Pr(B)$ 被加了 k 次. 除此之外，因 B 是事件 $A_1, \cdots, A_k$ 中每两个事件交

集的子集，那么在等式(1.7.2)的右端第二个和项中 Pr(B) 被加了 $\binom{k}{2}$ 次. 因此要将

41

Pr(B)减去 $\binom{k}{2}$ 次. 同样地，在第三个和项中，Pr(B) 被加了 $\binom{k}{3}$ 次. 这样继续下去，可知在等式(1.7.2)右端 Pr(B) 被重复使用的次数为：

$$k-\binom{k}{2}+\binom{k}{3}-\binom{k}{4}+\cdots+(-1)^{k+1}\binom{k}{k}.$$

可以用二项式定理证明这个数是 1.

由于同样的理由，对于属于 $A_1,\cdots,A_n$ 中任意给定的 k 个事件结果的子集，可以得到同样的结果. 因为 $k(1\leqslant k\leqslant n)$ 是任意的，所以从这个结论可以得出：$\bigcup_{i=1}^{n}A_i$ 中每个不相交子集的概率在等式(1.7.2)中刚好被计算了一次. ■

可将等式(1.7.2)的计算过程简述为：首先求 n 个单个事件的概率和，再减去任意两个事件交的概率的和；这一步会包括 $\binom{n}{2}$ 个不同的事件对. 第三步，加上所有任意三个事件交的概率的和；这一步共有 $\binom{n}{3}$ 个不同的事件交. 第四步，减去所有任意四个事件交的概率的和；这一步共有 $\binom{n}{4}$ 个不同的事件交. 继续这样的方法直到最后，根据 n 是奇数还是偶数来确定加上或减去所有 n 个事件的交的概率.

1.7.3 扩展研究：匹配问题

把一副有 n 张不同的牌排成一行，接着把另一副同样的牌洗牌后排成一行，放在原来那副牌的上面. 要求的是在两副对应的牌之间至少有一张相匹配的概率 p_n. 该问题可以有不同的有趣的表述内容. 比如，设一个人打印了 n 封信，并在 n 个信封上打印相应的地址，接着随意地将这 n 封信放进 n 个信封中. 要求至少有一封信刚好放进正确的信封的概率 p_n. 另一个例子是，随机地将 n 位著名影星的照片和他们幼年时候的照片配对，求至少有一对照片相匹配的概率 p_n.

下面我们以装信来讨论匹配的问题. 记 A_i 为把第 $i(i=1,\cdots,n)$ 封信放进正确的信封这一事件，利用等式(1.7.2)来计算 $p_n=\Pr\left(\bigcup_{i=1}^{n}A_i\right)$ 的值. 因为是随意把信装进信封的，那么把任意一封信放进正确的信封的概率 $\Pr(A_i)$ 是 $1/n$. 因此在等式(1.7.2)右边的第一项和的值是：

42

$$\sum_{i=1}^{n}\Pr(A_i)=n\cdot\frac{1}{n}=1.$$

进而，因为可把第一封信放进 n 个信封的任一个，那么第二封信可放进剩下的 $n-1$个信封中的任一个，则把第一封信和第二封信都放进正确的信封的概率 $\Pr(A_1A_2)$是 $1/[n(n-1)]$. 类似地，把任意第 i 封信和第 $j(i\neq j)$ 封信都放进正确的信封的概率 $\Pr(A_iA_j)$是 $1/[n(n-1)]$. 因此，等式(1.7.2)右边的第二项和的值是：

$$\sum_{i<j}\Pr(A_iA_j) = \binom{n}{2}\frac{1}{n(n-1)} = \frac{1}{2!}.$$

用类似的推理方法，可算出把任意3 封信 i、j 和 $k(i<j<k)$ 放进正确的信封的概率 $\Pr(A_iA_jA_k)$是 $1/[n(n-1)(n-2)]$. 因此，第三项和的值是：

$$\sum_{i<j<k}\Pr(A_iA_jA_k) = \binom{n}{3}\frac{1}{n(n-1)(n-2)} = \frac{1}{3!}.$$

持续这个过程，直到算得把所有 n 封信都正确放进各自信封的概率 $\Pr(A_1A_2\cdots A_n)$是 $1/n!$. 从等式(1.7.2)可得至少把一封信放进正确信封的概率 p_n 是：

$$p_n = 1-\frac{1}{2!}+\frac{1}{3!}-\frac{1}{4!}+\cdots+(-1)^{n+1}\frac{1}{n!}. \tag{1.7.3}$$

该概率有如下有趣的特征. 当 $n\to\infty$ 的时候，p_n 的值趋向于下式的极限：

$$\lim_{n\to\infty}p_n = 1-\frac{1}{2!}+\frac{1}{3!}-\frac{1}{4!}+\cdots.$$

在初等微积分的书中证明了该方程右端的无限序列的级数和等于 $1-(1/e)$，这里的 $e=2.71828\cdots$. 因此，$1-(1/e)=0.63212\cdots$. 即当 n 充分大时，至少有一封信被放进正确的信封的概率 p_n 的值接近于 0.63212.

如等式(1.7.3)所示，p_n 的准确值会随着 n 的增加形成一个波动序列. 当 n 以偶数2,4,6,…增加时，p_n 的值会逐渐增大到极限值0.63212；当 n 以奇数3,5,7,…增加时，p_n 的值会逐渐减小到同样的极限值.

p_n 的值收敛得很快. 事实上，当 $n=7$ 时，p_7 的值已经可以精确到小数点后第四位了. 因此，不管是把 7 封信随意地放进 7 个信封，还是把 7 百万封信随意地放进 7 百万个信封，至少有一封信放进正确信封的概率都是 0.6321.

43

1.7.4 小结

我们将任意两个事件和的概率公式推广到任意有限多个事件的和. 顺便说一下，在某些情况下，利用公式 $(A_1\cup\cdots\cup A_n)^c=A_1^c\cdots A_n^c$ 的事实通过计算 $1-\Pr(A_1^c\cdots A_n^c)$ 得到 $\Pr(A_1\cup\cdots\cup A_n)$更容易一些.

习题

1. 一副牌有 52 张，其中有 4 张 A，随机地把这副牌分给 3 个玩家，每人 5 张. 求至少有一个人手中的 5 张牌刚好有 2 张是 A 的概率.
2. 在某城中发行 A、B 和 C 三种报纸. 假设城市中有 60% 的家庭订阅报纸 A，40% 的家庭

订阅报纸B，30%的家庭订阅报纸C．再设有20%的家庭同时订阅报纸A和B，有10%的家庭同时订阅报纸A和C，有20%的家庭同时订阅报纸B和C，5%的家庭同时订阅三种报纸．求该市至少订阅三种报纸中的一种的家庭所占的比例？

3. 条件同习题2，求该市只订阅三种报纸中的一种的家庭所占的比例．
4. 把三张唱片从各自的套中取出，播放完毕后随机地把它们放回三个空套中，求至少有一张与原来套子匹配的概率．
5. 四位客人到达一家旅馆，并寄存了他们的帽子．离开时随机地领走帽子．求没有一位客人拿到自己帽子的概率．
6. 一个盒子装了30个红球，30个白球和30个蓝球．从中随机地不放回地取出10个球，求至少有一种颜色的球没有被拿到的概率p．
7. 一个学校的乐队中有一年级的学生10名、二年级的学生20名、三年级的学生30名和四年级的学生40名．若从这个乐队中任意地挑出15名学生，求每个年级至少有一名学生被挑到的概率．提示：先求四个年级中至少有一个年级不在这个选择之列的概率．
8. 随意地把n封信放进n个信封，求恰好有$n-1$封信被放进正确的信封的概率．
9. 设随意地把n封信放进n个信封，记q_n表示没有一封信被放入正确的信封的概率．求对n的下列四个值中哪个q_n最大：$n=10$，$n=21$，$n=53$，$n=300$．
10. 随意地把3封信放进3个信封，求恰好有1封信被放进正确的信封的概率．
11. 有10张卡片，其中红绿各5张．随机地把它们放进10个信封，信封的颜色也是红绿各5个，求刚好有$x(x=0,1,\cdots,10)$个信封里装的卡片与信封颜色匹配的概率．
12. 设A_1，A_2，…是满足$A_1\subset A_2\subset\cdots$的无限事件序列．证明：

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right)=\lim_{n\to\infty}\Pr(A_n).$$

提示：定义序列$B_1,B_2,\cdots$，要证

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right)=\lim_{n\to\infty}\Pr\left(\bigcup_{i=1}^{n} B_i\right)=\lim_{n\to\infty}\Pr(A_n).$$

13. 设$A_1,A_2,\cdots$是满足$A_1\supset A_2\supset\cdots$的无限事件序列．证明：

$$\Pr\left(\bigcap_{i=1}^{\infty} A_i\right)=\lim_{n\to\infty}\Pr(A_n).$$

44

提示：考虑序列$A_1^c,A_2^c,\cdots$，应用习题12的结果．

第2章 条件概率

2.1 条件概率的定义

统计推断中概率的一个主要应用是观察到某些事件后对概率进行修正，已知事件 B 发生后事件 A 被修正过的概率就是在条件 B 下事件 A 的条件概率.

设进行的试验的所有可能结果构成的样本空间为 S，S 中所有事件都有确定的概率. 现在来研究当已知另外某个事件 B 发生后事件 A 的概率将如何改变，称 A 的这个新概率为在事件 B 发生的条件下事件 A 发生的条件概率(conditional probability of the event A given the event B has occurred)，记这个条件概率为 $\Pr(A|B)$，为方便起见，这个记号可简单读作在条件 B 下事件 A 的条件概率.

若已知事件 B 已经发生，则试验的结果是包含在 B 中的一个事件. 因此，要计算 A 将发生的概率，必须考虑出现在 B 中并导致 A 发生的那些事件的集合. 如图 2-1 所示，该集合确切地讲即为集合 AB. 于是自然地定义条件概率 $\Pr(A|B)$ 为概率 $\Pr(AB)$ 在总概率 $\Pr(B)$ 中所占的比例. 这种想法引出了以下的定义：若 A，B 为两个事件，且 $\Pr(B)>0$，则有

$$\Pr(A|B)=\frac{\Pr(AB)}{\Pr(B)}. \tag{2.1.1}$$

当 $\Pr(B)=0$ 时，条件概率 $\Pr(A|B)$ 没有定义.

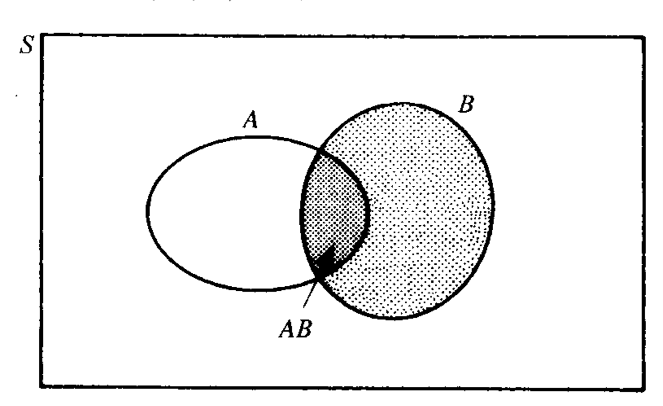


图 2-1 事件 B 中也属于事件 A 的结果

根据 1.2 节中概率的频率解释，条件概率 $\Pr(A|B)$ 有一个简单的意义. 按照这个解释，如果大量重复一个试验过程，则事件 B 发生的次数在重复试验总次数中所占的比例约为 $\Pr(B)$，事件 A 和 B 同时发生在整个重复过程中所占比例约为 $\Pr(AB)$. 因此，在事件 B 重复发生的试验中，事件 A 也发生的次数所占

的比例近似等于

$$\Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)}.$$

例2.1.1 掷骰子 设抛掷两个骰子并观察到所得两个点数之和 T 为奇数，求 T 小于8的概率.

如果用 A 表示事件 $T<8$，B 表示 T 为奇数的事件，则 AB 表示 $T=i$ ，$i=3,5,7$ 的事件.

由1.5节末给出的两个骰子的概率，可计算出 $\Pr(AB)$ 和 $\Pr(B)$ 如下：

$$\Pr(AB) = \frac{2}{36}+\frac{4}{36}+\frac{6}{36}=\frac{12}{36}=\frac{1}{3},$$

$$\Pr(B) = \frac{2}{36}+\frac{4}{36}+\frac{6}{36}+\frac{4}{36}+\frac{2}{36}=\frac{18}{36}=\frac{1}{2}.$$

于是 $$\Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)} = \frac{2}{3}.$$ ◀

例2.1.2 反复掷骰子 反复抛掷两个骰子，观察每次抛掷所得两个点数之和 T. 求 $T=7$ 先于 $T=8$ 出现的概率 p.

可如下直接计算欲求的概率 p：设样本空间 S 包含所有只要 $T=7$ 或 $T=8$ 就结束的抛掷结果的序列，然后求当观察到和值 $T=7$ 时就结束抛掷的所有序列的概率之和.

50

实际上，本例有一个更简单的方法. 考虑抛掷两个骰子的简单试验. 若重复试验，直至和 $T=7$ 或 $T=8$ 出现时就结束，效果就是限制试验的结果为这两个值之一. 因此，可重新将问题表述为：若给定试验的结果为 $T=7$ 或 $T=8$，求结果为 $T=7$ 的概率 P.

以 A 表示事件 $T=7$，B 表示 T 的取值为7或8的事件，则 $AB=A$，且

$$p = \Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)}.$$

由例1.5.4中掷两个骰子的概率得，$\Pr(A)=6/36$ 且 $\Pr(B)=(6/36)+(5/36)=11/36$. 于是 $p=6/11$. ◀

例2.1.3 诊疗试验 通常患有忧郁症的病人二至三年内会反复发作. Prien et al. (1984)研究了忧郁症的三种疗法：丙咪嗪，锂碳酸钙和联合疗法. 按照这类研究(称为诊疗试验)的惯例，将有一群病人服用安慰剂. (安慰剂是指既没有益处也没有害处的疗法. 让一些病人服用安慰剂是为了不让他们知道自己并没有接受其他任何一种治疗. 其余的病人也没有人会知道自己接受的是哪一种疗法或安慰剂.)本例中，考虑150名得过一段时期被称为“非狂躁型失常”类型的忧郁症的病人参加试验. 将他们分为四组治疗(三种疗法加上安慰剂)并持续观察有多少人发作. 表2-1综述了试验的结果. 若从该试验中随机选择一个病人并得知该病人接受了安慰剂，那么该病人复发的条件概率是多少? 设 B 表示病人接受安慰剂的事件，A 表示病人

复发的事件. 可直接从表中计算出 $\Pr(B)=34/150$ 且 $\Pr(AB)=24/150$，则 $\Pr(A|B)=24/34=0.706$. 另一方面，若发现随机选择的病人接受了锂(称为事件 C)，则有 $\Pr(C)=38/150$, $\Pr(AC)=13/150$ 和 $\Pr(A|C)=13/38=0.342$. 可知病人接受的疗法会影响复发的概率. 在第9章，将以更精确的方法来研究产生了多大的差异. ◀

表 2-1 忧郁症诊疗研究的结果

效果	治疗分组				总计
	丙咪嗪	锂	联合疗法	安慰剂	
复发	18	13	22	24	77
不复发	22	25	16	10	73
总计	40	38	38	34	150

51

2.1.1 条件概率的乘法法则

在一些试验中可以相对容易地直接计算某些条件概率. 在这些情况下，可以根据下面两个等式之一来计算两个事件同时发生的概率：

$$\Pr(AB)=\Pr(B)\Pr(A|B)$$

或

$$\Pr(AB)=\Pr(A)\Pr(B|A). \tag{2.1.2}$$

这些等式是从等式(2.1.1)和类似于 $\Pr(B|A)$ 的定义直接推出的.

例 2.1.4 选择两个球 假定从一个装有 r 个红球和 b 个蓝球的箱子里随机不放回地取出两个球. 求取出的第一个球是红球且第二个球是蓝球的概率 p.

以 A 表示取出的第一个球是红球的事件，B 表示取出的第二个球是蓝球的事件. 显然有 $\Pr(A)=r/(r+b)$. 进一步，如果事件 A 发生，则第一次取出的红球将从箱子里移走. 因此，第二次取到蓝球的概率为

$$\Pr(B|A)=\frac{b}{r+b-1}.$$

由此

$$\Pr(AB)=\frac{r}{r+b}\cdot\frac{b}{r+b-1}.$$ ◀

将这个原理扩展到有限多个事件. 如下面定理所示：

定理 2.1.1 假定 $A_1,A_2,\cdots,A_n$ 表示事件，且 $\Pr(B)>0,\Pr(A_1A_2\cdots A_{n-1})>0$. 则有

$$\Pr(A_1A_2\cdots A_n)=\Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1A_2)\cdots\Pr(A_n|A_1A_2\cdots A_{n-1}). \tag{2.1.3}$$

证明： 等式(2.1.3)右端概率的乘积等于

$$\Pr(A_1)\cdot\frac{\Pr(A_1A_2)}{\Pr(A_1)}\cdot\frac{\Pr(A_1A_2A_3)}{\Pr(A_1A_2)}\cdots\frac{\Pr(A_1A_2\cdots A_n)}{\Pr(A_1A_2\cdots A_{n-1})},$$

因为 $\Pr(A_1A_2\cdots A_n)>0$，这个乘积中每个分母都为正. 乘积中除最后一项的分子

$\Pr(A_1A_2\cdots A_n)$外其余所有的项可相互抵消，即得等式(2.1.3)左端. ■

例2.1.5 选取四个球 假定从一个装有r个红球和b个蓝球($r\geqslant 2, b\geqslant 2$)的箱子中随机且不放回地选取四个球，一次取一个. 求取出的球的次序为红、蓝、红、蓝的概率.

52

设R_j表示第j次取出的是红球的事件，B_j表示第j次取出的是蓝球($j=1,\cdots,4$)的事件，则

$$\Pr(R_1B_2R_3B_4)=\Pr(R_1)\Pr(B_2|R_1)\Pr(R_3|R_1B_2)\Pr(B_4|R_1B_2R_3)$$
$$=\frac{r}{r+b}\cdot\frac{b}{r+b-1}\cdot\frac{r-1}{r+b-2}\cdot\frac{b-1}{r+b-3}.$$ ◀

注：条件概率和概率的用法很类似. 在本书中的所有情况下，可以证明在给定一个事件$B[\Pr(B)>0]$发生时每一个结果都有条件概率的形式. 只须用事件B发生的条件概率来代替所有的概率，用给定事件CB发生的条件概率代替在已知其他事件C发生下所有的条件概率即可. 例如，定理2.1.2就是定理2.1.1乘法法则的条件概率形式. 尽管给出了定理2.1.2的证明，我们将不再给出对所有此类条件概率定理的证明，因为它们的证明过程与无条件情形下的证明过程是完全类似的.

定理2.1.2 若$A_1,A_2,\cdots,A_n$，B为事件，且$\Pr(A_1A_2\cdots A_{n-1}|B)>0$. 则有

$$\Pr(A_1A_2\cdots A_n|B)=\Pr(A_1|B)\Pr(A_2|A_1B)\cdots\Pr(A_n|A_1A_2\cdots A_{n-1}B). \quad (2.1.4)$$

证明：等式(2.1.4)右端的概率的乘积等于

$$\frac{\Pr(A_1B)}{\Pr(B)}\cdot\frac{\Pr(A_1A_2B)}{\Pr(A_1B)}\cdots\frac{\Pr(A_1A_2\cdots A_nB)}{\Pr(A_1A_2\cdots A_{n-1}B)}.$$

因为$\Pr(A_1A_2\cdots A_{n-1}|B)>0$，乘积中每个分母都为正. 乘积中除第一个分母与最后一个分子外的每一项都相互抵消，得到$\dfrac{\Pr(A_1A_2\cdots A_nB)}{\Pr(B)}$，即为等式(2.1.4)的左端. ■

2.1.2 扩展研究：掷双骰子游戏

我们借助一个流行的称为掷双骰子的赌博游戏的讨论来结束本节. 这个游戏的一种玩法如下：一人抛掷两个骰子，观察出现的两个点数之和. 如果第一次掷出的和为7或11，他立刻就赢了. 若第一次掷出的和为2，3或12，他立刻就输了. 若第一次掷出的和是4，5，6，8，9或10，则重新抛掷骰子直至和为7或原来的值. 如果第二次在得到7之前先得到的是原来的值，这个人就赢了. 否则，若第二次先得到7而不是原来的值，那么这个人就输了.

现计算掷骰子的人赢的概率p. 第一次抛掷得到的和为7或11的概率π_0为

53

$$\pi_0=\frac{6}{36}+\frac{2}{36}=\frac{8}{36}=\frac{2}{9}.$$

若第一次得到的和是4，该人赢的概率q_4等于在得到和7之前再次得到和为4的条件概率. 如例2.1.2所述，这个概率相当于结果必为4或7时得到和4的概率.

因此

$$q_4 = \frac{3/36}{(3/36)+(6/36)} = \frac{1}{3}.$$

因为第一次抛掷得到的和为 4 的概率 $p_4 = 3/36 = 1/12$，于是，第一次掷得到的和为 4 且赢了的概率

$$\pi_4 = p_4 q_4 = \frac{1}{12}\cdot\frac{1}{3} = \frac{1}{36}.$$

类似地，第一次掷得和为 10 的概率 $p_{10} = 1/12$，第一次掷得和为 10 时且赢得游戏的概率为 $q_{10} = 1/3$. 因此，第一次掷得和为 10 且赢了的概率为

$$\pi_{10} = p_{10} q_{10} = \frac{1}{12}\cdot\frac{1}{3} = \frac{1}{36}.$$

同理对 $i = 5,6,8,9$ 时定义 p_i，q_i 和π_i 且计算得

$$\begin{aligned}
\pi_5 &= p_5 q_5 = \frac{4}{36}\cdot\frac{4/36}{(4/36)+(6/36)} = \frac{2}{45},\\
\pi_9 &= p_9 q_9 = \frac{2}{45},\\
\pi_6 &= p_6 q_6 = \frac{5}{36}\cdot\frac{5/36}{(5/36)+(6/36)} = \frac{25}{396},\\
\pi_8 &= p_8 q_8 = \frac{25}{396}.
\end{aligned}$$

因为该人将会以某种方式赢得游戏的概率 p 是以上所有计算得到的概率之和，可得

$$\begin{aligned}
p &= \pi_0 + (\pi_4 + \pi_{10}) + (\pi_5 + \pi_9) + (\pi_6 + \pi_8)\\
&= \frac{2}{9} + 2\cdot\frac{1}{36} + 2\cdot\frac{2}{45} + 2\cdot\frac{25}{396} = \frac{244}{495} = 0.493.
\end{aligned}$$

于是，掷双骰子游戏中赢得游戏的概率略小于 1/2.

54

2.1.3 小结

已知事件 $B[\Pr(B)>0]$发生后事件 A 的修正概率是在条件 B 下 A 的条件概率，记作 $\Pr(A|B)$，用 $\Pr(AB)/\Pr(B)$计算. 通常容易直接得到条件概率 $\Pr(A|B)$. 在这种情况下，用条件概率的乘法法则来计算 $\Pr(AB)=\Pr(B)\ \Pr(A|B)$. 所有的概率结果在事件 $B[\Pr(B)>0]$发生的条件下都有条件概率形式，只要将概率全部换成除原有条件外还依赖事件 B 的条件概率. 例如，条件概率下的乘法定理变为 $\Pr(A_1A_2|B) = \Pr(A_1|B)\Pr(A_2|A_1B)$.

习题

1. 若 $A\subset B$ 且有 $\Pr(B)>0$. 求 $\Pr(A|B)$的值.
2. 若 A 和 B 为不相交事件且 $\Pr(B)>0$，求 $\Pr(A|B)$的值.
3. 若 S 为一试验的样本空间，A 为空间中任意事件，求 $\Pr(A|S)$的值.

4. 一顾客每次购买牙膏都选择品牌 A 或 B. 假定初次购买后，以后他每一次购买时选择上次购买的品牌的概率为1/3，换成另一品牌的概率为2/3. 若第一次购买时该顾客选择 A 或 B 的概率相等，求他第一次和第二次都购买 A 牌牙膏且第三次和第四次都购买 B 牌牙膏的概率.
5. 一个箱子装有 r 个红球和 b 个蓝球. 从箱中随机取出一球并观察其颜色，然后将这个球和另外 k 个与其相同颜色的球放入箱中. 随机取出第二个球，观察颜色，并将它和另外 k 个同颜色球放入箱子. 每次取出一个球，重复进行这个过程. 若取出四个球，求前三个是红球且第四个是蓝球的概率.
6. 一个箱子内有三张卡片. 一张卡片两面都是红色，一张两面都是绿色，还有一张卡片一面是红色另一面是绿色. 从箱子中随机选出一张卡片，观察其一面的颜色. 若该面是绿色，求另一面也是绿色的概率.
7. 重新考虑1.7中习题2的条件. 若从城市中订阅 A 报纸的家庭中随机选出一家，求该家庭同时订阅 B 报纸的概率.
8. 重新考虑1.7中习题2的条件. 若从该城市至少订阅了报纸 A，B，C 中一种的家庭中随机选出一家，求该家庭订阅 A 报纸的概率.
9. 假定一个箱子里装有一个蓝色卡片和四个分别标记为 A、B、C、D 的红色卡片. 从箱子中的五个卡片中随机且无放回地选出两个.
   a. 若已知卡片 A 被选中，求选出的两个卡片都是红色的概率.
   b. 若已知至少选出一个红色的卡片，求两个卡片都是红色的概率.
10. 考虑以下掷骰子游戏的另一种玩法：一人掷两个骰子. 如果第一次掷出的点和为7或11，他立刻就赢得游戏. 若第一次掷出的点和为2，3或12，他立刻就输了. 若第一次掷出的点和是4，5，6，8，9或10，则重新抛掷骰子直至和为7或11或原来的值. 如果第二次在得到7或11之前先得到的是原来的值，这个人就赢了. 否则若第二次在得到原来的值之前得到7或11，那么这个人就输了. 求这个人赢得比赛的概率.
11. 对任意三个事件 A，B 和 D，且 $\Pr(D)>0$，证明：

55

$$\Pr(A\cup B\mid D)=\Pr(A\mid D)+\Pr(B\mid D)-\Pr(AB\mid D).$$

2.2 独立事件

如果 B 的发生不改变 A 的概率，则称 A 和 B 是独立的. 有许多 A 和 B 不独立的情形，但是，如果其他的事件 C 已经发生，A、B 就独立，这种情况下称给定 C 下 A 和 B 条件独立.

2.2.1 独立的定义

已知事件 B 发生的条件下事件 A 的条件概率是当知道事件 B 发生事件 A 的修正概率. 但是，有时候即使已经知道 B 发生，但不需要修正 A 的概率. 例如，先投一个硬币然后掷一个骰子，记 A 为骰子点数为3的事件，记 B 是硬币正面朝上的事件. 如果投硬币和掷骰子是互相孤立的，可以十分确定地给出 $\Pr(A\mid B)=\Pr(A)=1/6$.

在这种情况下，我们称 A 和 B 是独立事件.

一般地，如果 $\Pr(B)>0$,可以将等式 $\Pr(A|B)=\Pr(A)$写为

$$\frac{\Pr(AB)}{\Pr(B)}=\Pr(A).$$

如果上面的等式两边同乘以 $\Pr(B)$，可以得到等式 $\Pr(AB)=\Pr(A)\Pr(B)$. 为了避免 $\Pr(B)>0$ 的限制条件，可把两个事件独立的数学定义表述如下：如果 $\Pr(AB)=\Pr(A)\Pr(B)$，那么两个事件 A 和 B 独立. 假设 $\Pr(A)>0$ 且 $\Pr(B)>0$，则由独立和条件概率的定义易得：A 和 B 独立的充要条件是 $\Pr(A|B)=\Pr(A)$ 且 $\Pr(B|A)=\Pr(B)$.

2.2.2 两个事件的独立

设事件 A 和 B 是独立的，因为两事件在物理上不相关，如果已知概率 $\Pr(A)$和 $\Pr(B)$，则可以用定义确定 $\Pr(AB)$的值.

例 2.2.1 机器运行 假设一个工厂里的两台机器 1 和 2 相互独立地运行. 事件 A 是机器 1 在给定的 8 小时内不运行，事件 B 是在同样时间内机器 2 不运行. 设 $\Pr(A)=1/3$,$\Pr(B)=1/4$. 求至少有一台机器在给定的时间内不运行的概率.

两台机器在给定的时间内不运行的概率 $\Pr(AB)$为

$$\Pr(AB)=\Pr(A)\Pr(B)=\left(\frac{1}{3}\right)\left(\frac{1}{4}\right)=\frac{1}{12}.$$

56

因此，至少有一台机器在给定的时间内不运行的概率 $\Pr(A\cup B)$是

$$\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(AB)=\frac{1}{3}+\frac{1}{4}-\frac{1}{12}=\frac{1}{2}.$$ ◀

下面的例子说明事件 A 和 B 虽然物理上有联系，但仍满足独立的定义.

例 2.2.2 掷骰子 假设掷一个均匀的骰子. 设 A 为得到偶数点数的事件，设 B 为点数是 1、2、3、4 中的任意一个数的事件. 下面说明事件 A 和 B 是独立的.

在这个例子中，$\Pr(A)=1/2$，$\Pr(B)=2/3$. 由于 AB 是得到 2 或者 4 的事件，$\Pr(AB)=1/3$. 因此，$\Pr(AB)=\Pr(A)\Pr(B)$. 这说明事件 A 和 B 是独立事件，尽管这两个事件的发生取决于掷同一个骰子的试验. ◀

在例 2.2.2 中也可以这样解释事件 A 和 B 独立：设一个人必须根据掷骰子结果是奇数还是偶数来下赌注，即，事件 A 是否发生. 因为可能的结果中有 3 个是奇数，其他 3 个是偶数，这个人在奇数上下赌注和在偶数上下赌注之间是无偏向性的.

再设掷骰子之后，但是在这个人知道结果和她决定在奇数结果上下赌注或者在偶数结果上下赌注之前，告诉她实际结果是 1、2、3、4 中的任意一个，即事件 B 已经发生. 这个人现在知道结果是 1、2、3 或者 4. 不过，因为这些数中 2 个是奇数，2 个是偶数，这个人对下赌注奇偶数仍然是无偏向性的. 换句话说，事件 B 发生的信息对试图确定事件 A 是否发生的这个人来说是没有帮助的.

在前面独立事件的讨论中，如果知道事件 A 和 B 独立，那么 A 发生或者不发生

和 B 发生或者不发生是没有关系的. 因此, 如果 A 和 B 满足独立事件的数学定义, 那么 A 和 B^c、A^c 和 B、A^c 和 B^c 也都是互相独立的事件. 在下面的定理中证明以上结论中的一个.

定理 2.2.1 如果事件 A 和 B 独立, 则事件 A 和 B^c 也独立.

57

证明: $\Pr(AB^c)=\Pr(A)-\Pr(AB)$ 是恒成立的.

进而, 因为 A 和 B 是独立事件

$$\Pr(AB)=\Pr(A)\Pr(B).$$

于是可得

$$\Pr(AB^c)=\Pr(A)-\Pr(A)\Pr(B)=\Pr(A)[1-\Pr(B)]=\Pr(A)\Pr(B^c),$$

所以, 事件 A 和 B^c 独立. ■

类似可证 A^c 和 B 也是独立的, A^c 和 B^c 独立的证明放在本节的习题 2 中.

2.2.3 几个事件的独立

可以将关于两个事件的讨论扩展到任意多个事件. 如果 k 个事件 $A_1,\cdots,A_k$ 独立的意义是它们物理上互不相关, 那么很自然假设 k 个事件将发生的概率 $\Pr(A_1\cap\cdots\cap A_k)$ 是乘积 $\Pr(A_1)\cdots\Pr(A_k)$. 进一步, 既然事件 $A_1,\cdots,A_k$ 互不相关, 这个乘积法则不仅对于 k 个事件的交, 而且对于它们中任意两个、任意三个、任意多个的交都成立. 这些思考导出了下面的定义: 称 k 个事件 $A_1,\cdots,A_k$ *互相独立*, 如果对这些事件的每个子集 $A_{i_1},\cdots,A_{i_j}(j=2,3,\cdots,k)$ 有

$$\Pr(A_{i_1}\cap\cdots\cap A_{i_j})=\Pr(A_{i_1})\cdots\Pr(A_{i_j}).$$

特别地, 为使三个事件 A、B、C 互相独立, 必须满足下面四个关系式:

$$\Pr(AB)=\Pr(A)\Pr(B),$$
$$\Pr(AC)=\Pr(A)\Pr(C),$$
$$\Pr(BC)=\Pr(B)\Pr(C), \tag{2.2.1}$$

和

$$\Pr(ABC)=\Pr(A)\Pr(B)\Pr(C). \tag{2.2.2}$$

满足等式(2.2.2); 但是不都满足(2.2.1)全部三个等式的情形是可能的. 另一方面, 下面的例子中, (2.2.1)中的三个等式都成立, 但是不满足等式(2.2.2).

例 2.2.3 两两独立 考虑一个试验的样本空间 S 包含四个结果 $\{s_1,s_2,s_3,s_4\}$, 假设每个结果的概率都是 1/4. 定义三个事件 A、B、C 如下:

58

$$A=\{s_1,s_2\},B=\{s_1,s_3\},\quad C=\{s_1,s_4\}.$$

那么 $AB=AC=BC=ABC=\{s_1\}$. 因此,

$$\Pr(A)=\Pr(B)=\Pr(C)=1/2,$$

且

$$\Pr(AB)=\Pr(AC)=\Pr(BC)=\Pr(ABC)=1/4.$$

这说明它满足式(2.2.1)中的三个等式, 但是不满足等式(2.2.2). 可以总结这个结果为事件 A、B、C 是两两独立的, 但是三个事件不互相独立. ◀

下面举出一些例子来说明独立的概念在解决概率问题中的功能和范围.

例 2.2.4 检查产品 设一台机器生产出废品的概率是 $p(0<p<1)$，生产出正品的概率为 $q=1-p$. 随机抽取和检查机器生产出来的 6 件产品，这 6 件产品的结果是互相独立的. 求 6 件产品中正好有 2 件是废品的概率.

可设样本空间 S 包括 6 件产品所有可能的排列，它们中的任一个可能是正品或者废品. 记 D_j 表示其中第 j 个产品是废品的事件，N_j 表示它是正品的事件. 因为 6 个不同产品的结果是互相独立的，所以生产出正品和废品的某个特定序列的概率是单个产品概率的乘积. 例如：

$$\Pr(N_1D_2N_3N_4D_5N_6)=\Pr(N_1)\Pr(D_2)\Pr(N_3)\Pr(N_4)\Pr(D_5)\Pr(N_6)=qpqqpq=p^2q^4.$$

可以看到 S 中其他包括 2 件废品和 4 件正品的特定序列的概率也是 p^2q^4. 因此，在样本空间的 6 件产品中正好有 2 件废品的概率是任一包括 2 件废品的特定序列的概率 p^2q^4 乘以所有这样可能的序列的数目. 因为有 2 件废品和 4 件正品不同的列数是 $\binom{6}{2}$，正好有 2 件废品的概率是 $\binom{6}{2}p^2q^4$. ◀

例 2.2.5 得到一件废品 条件同例 2.2.4，求在样本中 6 件产品至少有 1 件产品是废品的概率. 因为不同产品的结果是相互独立的，6 件产品都是正品的概率为 q^6. 因此，至少有 1 件产品是废品的概率是 $1-q^6$. ◀

例 2.2.6 抛掷一个硬币直到出现正面 抛掷一个硬币直到第一次出现正面，假设抛掷的结果是独立的. 求正好需要 n 次投掷的概率 p_n.

59

所求的概率等于连续得到 $n-1$ 次反面，在接下来的一次抛掷中得到正面的概率. 因为抛掷结果是独立的，所求事件的概率就是 $p_n=(1/2)^n$.

迟早得到正面的概率（或等价地，永远不会得到反面的概率）是

$$\sum_{n=1}^{\infty}P_n=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots=1$$

因为概率 p_n 的和是 1，表明得到无限多次反面而没有得到正面的概率一定是 0. ◀

例 2.2.7 一次检查一件产品 再次考虑机器生产出废品的概率为 p，生产出正品的概率为 $q=1-p$. 设随机抽取机器生产出来的产品，一次检查一件直到恰好得到五件废品. 求必须抽取 $n(n\geqslant5)$ 件产品恰好得到五件废品的概率 p_n.

检查中的第 n 件产品是第五件废品的充要条件是前 $n-1$ 件产品中恰好有四件废品且第 n 件产品是废品. 根据例 2.2.4 中相似的理由，在前 $n-1$ 件产品中恰好有四件废品和 $n-5$ 件正品的概率为 $\binom{n-1}{4}p^4q^{n-5}$. 第 n 件产品是废品的概率是 p. 因为第一个事件涉及的结果仅和前 $n-1$ 件产品有关，第二个事件涉及的结果仅和第 n 件产品有关，这两个事件独立. 所以，两个事件都发生的概率等于它们的概率乘积. 即

$$p_n = \binom{n-1}{4} p^5 q^{n-5}.$$ ◀

例 2.2.8 People v. Collins Finkelstein and Levin(1990)描述了一个犯罪案例，加利福尼亚高级法院撤销了裁决，它部分地涉及条件期望和独立的概率计算. 这个案例涉及抢钱包，目击者称看见一个金黄色头发、马尾发型的年轻女子是坐在一辆由留着胡子的黑人男子开着的黄色小汽车里逃离现场的. 犯罪活动发生后几天逮捕了符合上述特征的一对夫妇，但是没有发现实际证据. 一位数学家计算了随机选取的一对夫妇具有上述特征的概率大约是 8.3×10^{-8} 或者 120 万分之一. 面对如此巨大的可能性和没有实际证据，陪审团认定被告一定是这对夫妇，宣判他们有罪. 高级法院认为应该计算一个更有用的概率，根据目击者的证词，有一对夫妇符合上面的描述. 在已经有一对夫妇符合描述的条件下，存在符合被告特征的另一对夫妇的条件概率是多少?

60

设从 n 对夫妇中随机挑选一对具有特定特征的夫妇的概率为 p，记 A 为全体夫妇中至少有一对具有这些特征的事件，记 B 为至少有两对具有这些特征的事件. 求 $\Pr(B|A)$.

因为 $B\subset A$，

有
$$\Pr(B|A)=\frac{\Pr(BA)}{\Pr(A)}=\frac{\Pr(B)}{\Pr(A)}.$$

通过把每个事件划分为更易处理的小事件来计算 $\Pr(B)$ 和 $\Pr(A)$. 设把 n 对夫妇编号为 1 到 n. 记 A_i 为编号为 i 的夫妇具有正被考虑的特征的事件，$(i=1,\cdots,n)$，记 C 为正好有一对夫妇具有这些特征的事件，那么

$$A=(A_1^cA_2^c\cdots A_n^c)^c,$$

$$C=A_1A_2^c\cdots A_n^c\cup A_1^cA_2A_3^c\cdots A_n^c\cup\cdots\cup A_1^c\cdots A_{n-1}^cA_n,$$

$$B=A\cap C^c.$$

假设 n 对夫妇之间是相互独立的，$\Pr(A^c)=(1-p)^n$，$\Pr(A)=1-(1-p)^n$.

并集 C 中的 n 个事件互不相交，并且每个概率为 $p(1-p)^{n-1}$，因此 $\Pr(C)=np(1-p)^{n-1}$.

因为 $A=B\cup C$，且 B 和 C 不相交，有

$$\Pr(B)=\Pr(A)-\Pr(C)=1-(1-p)^n-np(1-p)^{n-1}$$

因此
$$\Pr(B|A)=\frac{1-(1-p)^n-np(1-p)^{n-1}}{1-(1-p)^n}. \qquad (2.2.3)$$

加利福尼亚高级法院的理由是：因为犯罪发生在人口众多的地方，n 是数以百万计的. 例如，设 $p=8.3\times10^{-8}$，$n=8\,000\,000$，则(2.2.3)的值为 0.2966. 这样一个概率表明存在一个合理的可能性，有另外一对夫妇符合目击者提供的描述. 当然，法院不知道 n 有多大，但是由于(2.2.3)的值相当大，事实上有足够的理由裁定，对被告是有罪的怀疑仍然合理的. ◀

独立和条件概率

两个具有正的概率的事件 A 和 B 独立的充要条件是 $\Pr(A|B)=\Pr(A)$. 对于多个独立事件也有类似的结果. 例如, 下面的定理, 可以根据独立的定义直接证明之.

定理 2.2.2 设事件 $A_1,\cdots,A_k$ 满足 $\Pr(A_1\cdots A_k)>0$, 则 $A_1,\cdots,A_k$ 互相独立的充要条件是对于 $\{1,\cdots,k\}$ 的任意两个不相交的子集 $\{i_1,\cdots,i_m\}$ 和 $\{j_1,\cdots,j_l\}$, 有

$$\Pr(A_{i_1}\cdots A_{i_m}|A_{j_1}\cdots A_{j_l})=\Pr(A_{i_1}\cdots A_{i_m}).$$ ■ 61

定理 2.2.2 表明 k 个事件互相独立的充要条件是在知道某些事件要发生的信息后并不改变其他事件组合的发生概率.

2.2.4 条件独立事件

条件概率和独立性组合成最复杂的数据收集模型之一. 也就是说, 在很多情况下不能说一些特定的事件互相独立, 因为我们认为在得知一些信息的情况下, 会提供其他事件可能发生的信息, 但是如果得知了事件发生的频率的信息后, 很可能就认为它们是独立的, 可以用本节较早出现的一个例子来描述这个模型.

例 2.2.9 检查产品 条件同例 2.2.4. 然而, 这次假设机器生产出废品的概率 p 是未知的. 为简单起见, 设 p 可能取 0.01 和 0.4 中的一个值, 第一种情况表明机器运转正常, 第二种情况表明机器需要维修. 记 B_1 代表 $p=0.01$ 的事件, 记 B_2 代表 $p=0.4$ 的事件. 如果已知事件 B_1 发生, 则可以假定事件 $D_1,D_2,\cdots$ 是互相独立的, 且对所有 i 有 $\Pr(D_i|B_1)=0.01$. 例如, 同例 2.2.4 和 2.2.7 中一样可以用 $p=0.01$ 做同样的计算. 设 A 为随机抽取 6 个产品中有两个是废品的事件, 则 $\Pr(A|B_1)=\binom{6}{2}0.01^2 0.99^4=1.44\times10^{-3}$. 类似地, 当 B_2 发生时, 可以认为 $D_1,D_2,\cdots$ 是互相独立的, 且有 $\Pr(D_i|B_2)=0.4$. 在这种情况下, $\Pr(A|B_2)=\binom{6}{2}0.4^2 0.6^4=0.311$. ◀

当然, 认为 p 一定同例 2.2.9 一样最多取两个不同的值是没有道理的. p 很容易取到第三个、第四个值等等. 实际上, 在第 3 章将学习如何处理 p 为 0 到 1 之间任意可能取值的情形. 这个简单例子的意义就是描述同例中的 B_1 或 B_2 一样, 在另一事件发生的条件下假设事件独立的概念.

条件独立的严格定义是: 在给定条件 B 下 $A_1,\cdots,A_k$ 互相独立的充分条件是对于这些事件中的任意 j 个子集 $A_{i_1},\cdots,A_{i_j}(j=2,3,\cdots,k)$,

$$\Pr(A_{i_1}\cap\cdots\cap A_{i_j}|B)=\Pr(A_{i_1}|B)\cdots\Pr(A_{i_j}|B).$$

此定义等价于把条件是全部概率修改为以 B 为条件的情况下的独立事件的定义. 在第 2.3 节学习了划分的概念后, 就会显示出条件独立性的作用. 作为注记, 即使假设在事件 B 下事件 $A_1,\cdots,A_k$ 条件独立, 不一定有在 B^c 下仍条件独立. 在例 2.2.9 中, 事件 $D_1,D_2,\cdots$ 在 B_1 和 $B_2=B_1^c$ 下都是条件独立的, 这是特殊情形. 62

回忆两事件 A_1 和 A_2 [$\Pr(A_1)>0$] 互相独立的充要条件是 $\Pr(A_2|A_1)=\Pr(A_2)$. 对于条件独立事件也有同样的结果.

定理 2.2.3 设事件 A_1、A_2 和 B 满足 $\Pr(A_1B)>0$，则 A_1 和 A_2 关于 B 条件独立的充要条件是 $\Pr(A_2|A_1B)=\Pr(A_2|B)$. ■

这个例子再次证明了我们前面提到的论断，可以证明在条件 B 下都有类似的结果. 读者可在习题22中证明这个定理.

2.2.5 扩展研究：收集者问题

假设随机地把 n 个球投进 r 个盒子($r\leqslant n$)，可以认为 n 次投球都是独立的，且 r 个盒子中的每一个也是等可能地接收球. 问题是确定每个盒子至少收到一个球的概率 p. 可将这个问题变化为如下收集者问题：设每盒泡泡糖都含有一幅篮球运动员的画片；共使用了 r 个不同运动员的画片；每幅运动员的画片都等可能地放进任一盒泡泡糖中；把画片装进盒子是彼此独立的. 现在的问题是求一个买了 n 盒泡泡糖的人得到全部 $r(n\geqslant r)$ 张不同画片的概率 p.

记 A_i 表示事件"第 i 个运动员的画片不在所有 n 盒中"，$i=1,\cdots,r$. 则 $\bigcup_{i=1}^{r}A_i$ 表示事件"至少有一名运动员的画片不在盒中". 应用等式(1.7.2)可得 $\Pr(\bigcup_{i=1}^{r}A_i)$.

因为 r 名运动员中每个人的画片都是等可能地被放进任意盒中的，则在任一特定盒中收集不到第 i 个运动员的画片的概率为 $(r-1)/r$. 因为装盒子是独立的，则在 n 个盒中的每一个都收集不到第 i 名运动员画片的概率是$[(r-1)/r]^n$. 因此，

$$\Pr(A_i)=\left(\frac{r-1}{r}\right)^n \quad i=1,\cdots,r.$$

现在考虑任意两名运动员 i 和 j，在任意一盒中第 i 名运动员的画片和第 j 名运动员的画片都不出现的概率是$(r-2)/r$. 因此，在 n 盒中的每一个中两张画片都不出现的概率为$[(r-2)/r]^n$，即

$$\Pr(A_iA_j)=\left(\frac{r-2}{r}\right)^n.$$

下面考虑任意三名运动员 i、j 和 k，有

63

$$\Pr(A_iA_jA_k)=\left(\frac{r-3}{r}\right)^n.$$

如此持续下去，最终可得所有 r 名运动员的画片都不出现在 n 个盒中的概率 $\Pr(A_1A_2\cdots A_r)$，当然，这个概率为0. 因此，由1.7节的等式(1.7.2)得

$$\begin{aligned}\Pr\left(\bigcup_{i=1}^{r}A_i\right)&=r\left(\frac{r-1}{r}\right)^n-\binom{r}{2}\left(\frac{r-2}{r}\right)^n+\cdots+(-1)^r\binom{r}{r-1}\left(\frac{1}{r}\right)^n\\&=\sum_{j=1}^{r-1}(-1)^{j+1}\binom{r}{j}\left(1-\frac{j}{r}\right)^n.\end{aligned}$$

因为收集到一整套 r 幅不同画片的概率 p 等于 $1-\Pr(\bigcup_{i=1}^{r} A_i)$，从前面的推导可得，$p$ 可表示为

$$p=\sum_{j=0}^{r-1}(-1)^j\binom{r}{j}\left(1-\frac{j}{r}\right)^n.$$

2.2.6 小结

一组事件是互相独立的充要条件是在获知其中一些事件发生的信息不改变其他事件的任意组合发生的概率. 等价地，一组事件互相独立的充要条件是每个子类的交的概率等于组成个体概率的乘积. 独立的概念有一个以另一事件为条件的版本. 一组事件在 B 条件下是互相独立的充要条件是以 B 为条件的任意子集的交的概率等于各个个体以 B 为条件的概率的乘积. 等价地在 B 条件下，一组事件条件独立的充要条件是在获知其中一些事件（包含 B）发生后却不改变其他事件任意组合在 B 为条件的发生概率. 在下一节介绍了划分后，条件独立性的全部作用会变得更加明显.

习题

1. 若 A 和 B 是独立事件，且 $\Pr(B)<1$，求 $\Pr(A^c\mid B^c)$的值.
2. 设 A 和 B 是独立事件，证明事件 A^c 和 B^c 也是互相独立的.
3. 假设事件 A 满足 $\Pr(A)=0$ 且 B 是任意一个其他事件，证明 A 和 B 是独立事件.
4. 一人先后掷两个均匀的骰子三次. 求三次掷的结果中每次出现的两个点数之和都是 7 的概率.
5. 设在某次飞行中，宇宙飞船所使用的控制系统失灵的概率是 0.001，设在飞船上另外安装了一个备用但完全独立的控制系统，在第一个系统失灵时使用. 求在这次飞行中，宇宙飞船处于原来的或者备用的系统控制下的概率.
6. 在一次博彩中卖出了 10 000 张彩票，在另一次卖出了 5000 张. 如果一人在每次博彩中都拥有 100 张彩票，求她至少赢得一次一等奖的概率.

64

7. A 和 B 两名学生都注册上一门课. 设学生 A 有 80% 的时间到课，学生 B 有 60% 的时间到课，两名学生缺席是互相独立的.
   a. 求一天中两名学生中至少有一名到课的概率是多少？
   b. 若某一天两名中至少有一名学生在上课，求这天 A 在上课的概率是多少？
8. 掷三个均匀的骰子，求出现的三个点数是一样的概率是多少？
9. 考虑一个试验，掷一个均匀的硬币直至第一次出现正面. 若这个试验进行了 3 次，求 3 次试验中每次都需要掷相同次数的概率.
10. 设某家庭中的任何一个孩子是蓝眼睛的概率是 1/4，这个特征遗传给每个孩子是相互独立的. 若该家有五个孩子，且至少有一名孩子是蓝眼睛，求至少有三个孩子是蓝眼睛的概率.
11. 考虑习题 10 中的这个家庭.
   a. 若已知此家庭中最小的孩子是蓝眼睛，求至少有三个孩子是蓝眼睛的概率.

b. 解释 a 中的答案为何与习题 10 的答案不同.

12. 设 A、B 和 C 是满足 $\Pr(A)=1/4$，$\Pr(B)=1/3$，$\Pr(C)=1/2$ 的三个独立事件.

    a. 求三个事件都没发生的概率.

    b. 求三个事件中刚好有一个发生的概率.

13. 设由一放射性物质释放的任何微粒会以 0.01 的概率穿透某种介质. 若释放出 10 个微粒，求正好有一个微粒穿透介质的概率.

14. 条件同习题 13，若释放出 10 个微粒，求至少有一个微粒穿透介质的概率.

15. 条件同习题 13，求需要释放多少微粒才能使至少有一个微粒穿透介质的概率不少于 0.8.

16. 在世界棒球联赛中，A 和 B 两队之间有一系列赛事，且先赢得四场比赛的队在世界联赛中胜出. 若在任意一场比赛中 A 队打败 B 队的概率是 1/3，求 A 队赢得世界联赛的概率.

17. A 和 B 两个男孩向一个目标投球. 设男孩 A 每次投中的概率是 1/3，男孩 B 每次投中的概率是 1/4. 再设男孩 A 先投且两男孩轮流投球. 求男孩 A 在第三次投掷时首次投中的概率.

18. 条件同习题 17，求男孩 A 比男孩 B 先投中的概率.

19. 一个盒子装有 20 个红球、30 个白球和 50 个蓝球. 从盒子中随机抽取一个球，然后再放回盒子，共进行 10 次；也就是说，在下次挑选前把已经选中的球再放回盒子中. 求在这 10 个选中的球中至少有一种颜色没有出现的概率.

20. 假设 $A_1,\cdots,A_k$ 是一列 k 个相互独立的事件. 记 $B_1,\cdots,B_k$ 是另一列 k 个事件，满足对于任意 $j(j=1,\cdots,k)$ 有 $B_j=A_j$ 或 $B_j=A_j^c$. 证明事件 $B_1,\cdots,B_k$ 也是相互独立的. 提示：由满足 $B_j=A_j^c$ 的事件 B_j 的数目归纳证明.

21. 证明定理 2.2.2.

    提示：“必要性”可由独立性定义直接得出.

    “充分性”可用独立性的定义并对 j 值用归纳法证明.

    记 $m=j-1$，记 $l=1$，$j_1=i_j$.

22. 证明定理 2.2.3.

23. 一名程序员计划编译一列 11 个类似的程序. 记 A_i 为事件“第 i 个程序编译成功” $i=1,\cdots,11$. 当编程工作容易时，程序员预期 80% 的程序会编译成功. 而当编程工作困难时，预期会有 40% 的程序编译成功. 记 B 为事件“编程工作容易”，则程序员认为事件 $A_1,\cdots,A_{11}$ 在事件 B 和 B^c 下都是条件独立的.

    a. 求在给定 B 条件下，11 个程序中有 8 个会编译通过的概率.

    b. 求在给定 B^c 条件下，11 个程序中有 8 个会编译通过的概率.

65

2.3 贝叶斯定理

假设考虑 k 个互不相容的将要发生的事件 $B_1,\cdots,B_k$，且在此基础上观察其他的事件 A. 如果对每个 i 知道概率 $\Pr(A\mid B_i)$，则在事件 A 给定的条件下，计算 B_i 的条件概率时，贝叶斯定理是一个非常有用的公式.

首先从一个典型的例子开始.

例 2.3.1 疾病检测 假设你正在马路上散步，注意到公共卫生部门正在为某一特定疾病做一个免费医疗检查．这个试验在下列意义下有 90% 的可靠性：如果一个人有这种疾病，试验结果呈阳性的概率是 0.9；反之，如果一个人没有这种疾病，试验结果呈阳性的概率是 0.1.

数据表明得这种疾病的机会只是万分之一．然而，由于这个试验既没有花费你任何代价，又快且无害，所以你决定做这个试验来检查自己是否有这种疾病．一些天以后，你了解到检查结果是阳性．问你有这种疾病的概率是多少？ ◀

例 2.3.1 的最后一个问题是设计贝叶斯定理问题的原型．已知有一些信息（试验的结果）和至少两个互不相容的不确定事件("有病"和"没有病")．如果所得到的信息告诉我们关于不确定事件的一些知识，那么就需要知道在已知信息的前提下，怎样修正事件的概率.

在回到例题之前，先给出贝叶斯定理的一般结构.

2.3.1 概率和分割

记 S 表示某个试验的样本空间，并且考虑 S 中 k 个互不相容的事件 $B_1,\cdots,B_k$，且满足 $\bigcup_{i=1}^{k} B_i = S$．即这些事件构成了 S 的一个分割．特别地，如果知道分割中哪一个事件发生了，选择组成分割的事件以便减少问题中的不确定性．

例 2.3.2 诊疗试验 考虑一个诊疗试验，正如例 2.1.3 中对忧郁症的研究，通过多次试验，每个病人都有两种可能结果，复发和不复发．将复发称为"失败"，不复发称为"成功"．现在只考虑使用丙咪嗪治疗组中的病人．假设在一个治疗组中的每个病人结果是成功的概率都为 p．不幸的是，在试验初期，不知道 p 的值．因为 p 值是本问题中不确定性的重要来源，所以将根据 p 可能的取值来分割样本空间．为简单起见，假设 p 取值于 $0,1/10,\cdots,9/10,1$ 这 11 个值中的一个(在第 3 章中，会处理对 p 更现实的模型)．让 B_j 表示 $p=(j-1)/10(j=1,\cdots,11)$ 的事件，则 $B_1,\cdots,B_{11}$ 构成一个分割． ◀

66

如果 k 个事件 $B_1,\cdots,B_k$ 构成样本空间 S 的一个分割，且 A 是样本空间 S 的另一个事件．所以事件 $B_1A,B_2A,\cdots,B_kA$ 将构成 A 的一个分割，如图 2-2 所示(A 与 $B_1,\cdots,B_5$ 的交集)．因此有：

$$A=(B_1A)\cup(B_2A)\cup\cdots\cup(B_kA).$$

又因为上式右边 k 个事件互不相容，则有：

$$\Pr(A)=\sum_{j=1}^{k}\Pr(B_jA).$$

最后，如果 $\Pr(B_j)>0, j=1,\cdots,k$，那么 $\Pr(B_jA)=\Pr(B_j)\Pr(A\mid B_j)$，从而有：

$$\Pr(A)=\sum_{j=1}^{k}\Pr(B_j)\Pr(A\mid B_j).$$

综上所述，便证明了下面的定理.

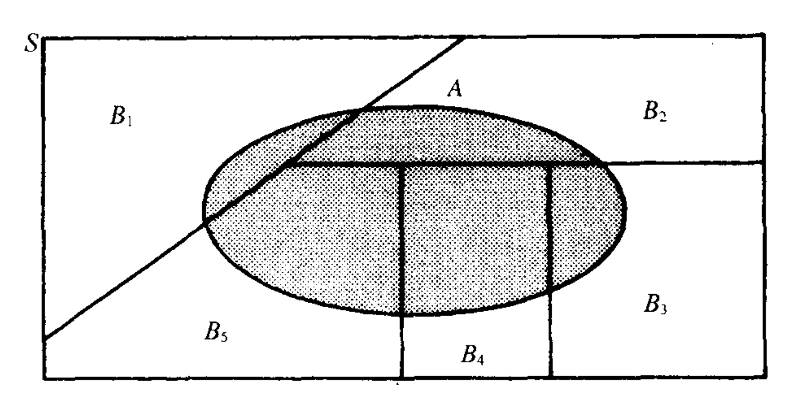


图2-2 A 与分割 $B_1,\cdots,B_5$ 的交集

定理 2.3.1 全概率公式 设事件 $B_1,\cdots,B_k$ 构成空间 S 的一个分割，且 $\Pr(B_j)>0$，$j=1,\cdots,k$. 则对于 S 中的任意事件 A，都有：

67

$$\Pr(A)=\sum_{j=1}^{k}\Pr(B_j)\Pr(A\mid B_j). \qquad \blacksquare$$

注：全概率公式的条件概率形式 在另一事件 C 的条件下，有类似的条件全概率公式，即

$$\Pr(A\mid C)=\sum_{j=1}^{k}\Pr(B_j\mid C)\Pr(A\mid B_jC). \tag{2.3.1}$$

在习题13中读者可证明这个公式.

例 2.3.3 选择螺钉 两个盒子都装有长螺钉和短螺钉. 假设一个盒子装有60个长螺钉和40个短螺钉，另一个盒子装有10个长螺钉和20个短螺钉，还假设随机地选取一个盒子然后随机地从该盒子中选取一个螺钉. 计算取到长螺钉的概率.

设 B_1 是选择第一个盒子的事件；B_2 是选择第二个盒子的事件；A 是取到一个长螺钉的事件. 那么，

$$\Pr(A)=\Pr(B_1)\Pr(A\mid B_1)+\Pr(B_2)\Pr(A\mid B_2).$$

因为是随机地选择盒子，所以 $\Pr(B_1)=\Pr(B_2)=1/2$. 进而从第一个盒子取到长螺钉的概率是 $\Pr(A\mid B_1)=60/100=3/5$. 从第二个盒子取到长螺钉的概率是 $\Pr(A\mid B_2)=10/30=1/3$. 因此，

$$\Pr(A)=\frac{1}{2}\cdot\frac{3}{5}+\frac{1}{2}\cdot\frac{1}{3}=\frac{7}{15}. \qquad \blacktriangleleft$$

例 2.3.4 诊疗试验 考虑如例2.1.3和2.3.2所描述的同一个诊疗试验. 设 E_1 是使用丙咪嗪治疗组中第一个病人治疗成功的事件. 在例2.3.2中，定义 $B_1,\cdots,B_{11}$ 使 $\Pr(E_1\mid B_j)=(j-1)/10$. 假设在开始试验之前，对每个 j 有 $\Pr(B_j)=1/11$，则

$$\Pr(E_1)=\sum_{j=1}^{11}\frac{1}{11}\cdot\frac{j-1}{10}=\frac{55}{110}=\frac{1}{2}. \tag{2.3.2}$$

其中第二个等式是根据 $\sum_{j=1}^{n}j=n(n+1)/2$ 得出来的. ———— $\blacktriangleleft$

例 2.3.5 获得一个高分 假设某人参加一个比赛，他的分数必是 $1,2,\cdots,50$ 这 50 个数中之一，并且他以等概率获得其中的每个分数. 第一次比赛，他的分数是 X. 然后继续比赛直到他获得另一个分数 Y，且 Y 满足 $Y\geqslant X$. 假设所有的比赛是相互独立的. 试求 $Y=50$ 的事件 A 的概率.

68

对每一个 $i=1,\cdots,50$，设 B_i 表示 $X=i$ 的事件. 在 B_i 发生的条件下，Y 以等概率取值于 $i,i+1,\cdots,50$ 中任何一个. 因为 Y 等可能地取 $(51-i)$ 值中的每一个，则

$$\Pr(A\mid B_i)=\Pr(Y=50\mid B_i)=\frac{1}{51-i}.$$

又因为 X 取这 50 个值中的每一个值的概率是 1/50，则对于每个 $i,\Pr(B_i)=1/50$，且

$$\Pr(A)=\sum_{i=1}^{50}\frac{1}{50}\cdot\frac{1}{51-i}=\frac{1}{50}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{50}\right)=0.0900. \qquad ◀$$

2.3.2 贝叶斯公式的陈述和证明

假设选择分割中的事件使得如果知道分割中的哪个事件发生了是降低不确定性的一个重要来源，如例 2.3.2. 在知道一些信息之后(如例 2.3.4 第一个病人复发)，分割中哪个事件发生了仍然是不确定的. 称之为贝叶斯定理的以下结果提供了一个计算在已知事件 A 的条件下，每一个分割事件 B_i 的条件概率的简单规则. 为计算上述条件概率，需要知道在已知每一个事件 B_i 下 A 的条件概率和每一个 B_i 的无条件概率.

定理 2.3.2 贝叶斯定理 设事件 $B_1,\cdots,B_k$ 构成样本空间 S 的一个分割，使得对 $j=1,\cdots,k$，B_j 满足 $\Pr(B_j)>0$，又设 A 是一个事件满足 $\Pr(A)>0$，则对 $i=1,\cdots,k$，有

$$\Pr(B_i\mid A)=\frac{\Pr(B_i)\Pr(A\mid B_i)}{\sum_{j=1}^{k}\Pr(B_j)\Pr(A\mid B_j)}. \tag{2.3.3}$$

证明: 由条件概率定义知，

$$\Pr(B_i\mid A)=\frac{\Pr(B_iA)}{\Pr(A)}.$$

根据条件概率(2.1.2)的乘法规则，等式(2.3.3)右边项的分子等于 $\Pr(B_iA)$. 根据定理 2.3.1，分母等于 $\Pr(A)$. ■

例 2.3.6 疾病检测 回到本节开始的例子. 我们刚接到对于这个疾病测试呈阳性的消息. 在例题 2.3.1 描述的意义下，试验有 90% 的可靠性. 确定试验呈阳性之后，想知道得这种疾病的概率. 一些读者可能认为这个概率应该为 0.9. 然而，这种认为完全忽视了在诊病之前有这种疾病的小概率为 0.0001. 记 B_1 为有这种疾病的事件，记 B_2 为没有这种疾病的事件. 同样，记 A 为试验结果呈阳性的事件. 则由贝叶斯定理

69

$$\Pr(B_1|A)=\frac{\Pr(A|B_1)\Pr(B_1)}{\Pr(A|B_1)\Pr(B_1)+\Pr(A|B_2)\Pr(B_2)}$$
$$=\frac{(0.9)(0.0001)}{(0.9)(0.0001)+(0.1)(0.9999)}=0.000\,90.$$

这样在给定试验结果的条件下，有这种疾病的条件概率大约是千分之一. 当然即使这个条件概率是十分小的，但这个条件概率大约是做检查之前概率的10倍. 这一结论另一种解释如下：实际上，每一万个人中仅有一个人有这种疾病，但是通过试验结果为阳性的条件下大约是每十个人中有一个人有这种疾病. 因此，呈阳性结果的人数大约是真正有这种疾病人数的1000倍. 也就是说，试验呈阳性结果的1000个人中，只有一个人真正有这种疾病. 通过这个例题不仅显示出贝叶斯定理的有用性，而且也显示出问题中考虑所有可得到的信息的重要性. ◀

例2.3.7 确定次品的来源 用三个不同的机器 M_1,M_2 和 M_3 生产一大批类似的部件. 假设20%的部件由 M_1 生产，30%的部件由 M_2 生产，50%的部件由 M_3 生产. 进一步假设机器 M_1 生产的部件是次品的概率为1%，机器 M_2 生产的部件是次品的概率为2%，机器 M_3 生产的部件是次品的概率为3%. 最后，假设从整批部件中随机地选择一个部件发现是次品. 试计算这个部件是由机器 M_2 生产的概率.

设 B_i 表示选中的部件是由机器 $M_i(i=1,2,3)$ 生产的事件，A 表示选中部件是次品的事件. 须计算条件概率 $\Pr(B_2|A)$.

设从整批部件中随机选择一个部件是由机器 M_i 生产的概率为 $\Pr(B_i)$，对 $i=1,2,3$ 有

$$\Pr(B_1)=0.2,\quad \Pr(B_2)=0.3,\quad \Pr(B_3)=0.5.$$

且由机器 M_i 生产的一个部件是次品的概率为 $\Pr(A|B_i)$，它们分别为

$$\Pr(A|B_1)=0.01,\Pr(A|B_2)=0.02,\Pr(A|B_3)=0.03.$$

根据贝叶斯定理，有

$$\Pr(B_2|A)=\frac{\Pr(B_2)\Pr(A|B_2)}{\sum_{j=1}^{3}\Pr(B_j)\Pr(A|B_j)}$$
$$=\frac{(0.3)(0.02)}{(0.2)(0.01)+(0.3)(0.02)+(0.5)(0.03)}=0.26.\ ◀$$

70

例2.3.8 确定基因类型 考虑有两个等位基因A和a，假设基因通过两种形式表现本身的特征(如头发颜色或血型). 具有基因类型AA和Aa的个体有相同的特征形式，而具有基因类型aa的个体有另一种特征形式，那么A和a分别称为显性和隐性. 特征的这两种形式称为表型. 具有基因类型AA和Aa的个体称为具有显性特征，具有基因类型aa的个体称为隐性特征. 在遗传学研究中，获得个体的表型信息是很普遍的，但是确定基因类型是十分困难的. 然而，通过观察父母和孩子的表

型特征能获得一些关于基因的信息.

假设等位基因 A 是显性的，个体基因配对是相互独立的，在群体中，出现基因类型 AA，Aa 和 aa 的概率分别是 1/4，1/2 和1/4. 我们去观察一个个体，并且不知道他的父母，观察这个个体的表型特征. 设 E 是被观察的个体有显性特征的事件. 我们希望揭示其父母可能有的基因类型. 在观察之前父母有 $B_1,\cdots,B_6$ 共6种可能的基因组合.

父母基因	(AA,AA)	(AA,Aa)	(AA,aa)	(Aa,Aa)	(Aa,aa)	(aa,aa)
事件名	B_1	B_2	B_3	B_4	B_5	B_6
B_i 的概率	1/16	1/4	1/8	1/4	1/4	1/16
$\Pr(E\mid B_i)$	1	1	1	3/4	1/2	0

假设父母基因因子是相互独立的，来计算 B_i 的概率. 例如，如果在父亲是 AA 和母亲是 aa(概率是 1/16)或者父亲是 aa 和母亲是 AA(概率是 1/16)的条件下，B_3 发生，从父母遗传给孩子的概率是 1/2，且双亲是相互独立的假设下，计算 $\Pr(E\mid B_i)$的值. 例如给定 B_4 条件下，事件 E 发生当且仅当孩子不能得到 2 个 a. 给定 B_4，从双亲得到 a 的概率是 1/4，所以 $\Pr(E\mid B_4)=3/4$.

现在计算 $\Pr(B_1\mid E)$和 $\Pr(B_5\mid E)$，其他计算留给读者. 对于这两个计算，贝叶斯定理的分母是一样的，即

$$\Pr(E)=\sum_{i=1}^{5}\Pr(B_i)\Pr(E\mid B_i)$$

$$=\frac{1}{16}\times 1+\frac{1}{4}\times 1+\frac{1}{8}\times 1+\frac{1}{4}\times\frac{3}{4}+\frac{1}{4}\times\frac{1}{2}+\frac{1}{16}\times 0=\frac{3}{4}.$$

应用贝叶斯定理,得:

$$\Pr(B_1\mid E)=\frac{\frac{1}{16}\times 1}{\frac{3}{4}}=\frac{1}{12},\quad \Pr(B_5\mid E)=\frac{\frac{1}{4}\times\frac{1}{2}}{\frac{3}{4}}=\frac{1}{6}.$$

◀ 71

注：条件贝叶斯定理的形式 在事件 C 条件下，有以下贝叶斯定理：

$$\Pr(B_i\mid AC)=\frac{\Pr(B_i\mid C)\Pr(A\mid B_iC)}{\sum_{j=1}^{k}\Pr(B_j\mid C)\Pr(A\mid B_jC)}\qquad(2.3.4)$$

2.3.3 先验和后验概率

在例2.3.7 中，常称 $\Pr(B_2)$为选中的部件由机器 M_2 生产的先验概率，因为 $\Pr(B_2)$是在选择部件且不知被选部件是不是次品之前事件的概率. 称 $\Pr(B_2\mid A)$为选中的部件由机器 M_2 生产的后验概率，因为 $\Pr(B_2\mid A)$是在已知选中次品之后的事件概率.

这样，在例2.3.7中，选中的部件是由机器 M_2 生产的先验概率是0.3. 选择一个部件且发现是次品之后，它是由机器 M_2 生产的后验概率是0.26. 因为这个后验概率小于该部件是由机器 M_2 生产的先验概率，则这个部件由其他机器生产的后验概率一定大于它是由其他机器生产的先验概率.（参见本节习题5和6）.

2.3.4 扩展研究：多阶段后验概率的计算

设想一个盒子里有一个均匀硬币和一个两面都是头像的硬币. 随机选取一个硬币进行投掷，得到一个头像. 我们要确定硬币是均匀硬币的概率.

令 B_1 为硬币是均匀硬币的事件，B_2 为硬币是有两个头像的事件，H_1 为投掷硬币得到一个头像的事件. 那么，由贝叶斯定理得到：

$$\begin{aligned}\Pr(B_1 \mid H_1) &= \frac{\Pr(B_1)\Pr(H_1 \mid B_1)}{\Pr(B_1)\Pr(H_1 \mid B_1) + \Pr(B_2)\Pr(H_1 \mid B_2)} \\ &= \frac{(1/2)(1/2)}{(1/2)(1/2) + (1/2)(1)} = \frac{1}{3}.\end{aligned} \tag{2.3.5}$$

从而，在第一次投掷后，硬币是均匀硬币的后验概率是1/3.

现在再次投掷相同的硬币，假定又得到一个头像. 假定这两次投掷在 B_1 和 B_2 的条件下是独立的. 有两种方法来确定硬币是均匀硬币的后验概率的新数值.

第一种方法是返回到试验的开始，再次假定先验概率是 $\Pr(B_1) = \Pr(B_2) = 1/2$. 72 用 H_1H_2 表示在硬币的两次投掷中都得到头像的事件. 我们要计算在观测到事件 H_1H_2 之后，硬币是均匀硬币的后验概率 $\Pr(B_1 \mid H_1H_2)$. 假设所有投掷在 B_1 条件下是独立的，这就意味着 $\Pr(H_1H_2 \mid B_1) = 1/2 \times 1/2 = 1/4$. 依据贝叶斯定理，

$$\begin{aligned}\Pr(B_1 \mid H_1H_2) &= \frac{\Pr(B_1)\Pr(H_1H_2 \mid B_1)}{\Pr(B_1)\Pr(H_1H_2 \mid B_1) + \Pr(B_2)\Pr(H_1H_2 \mid B_2)} \\ &= \frac{(1/2)(1/4)}{(1/2)(1/4) + (1/2)(1)} = \frac{1}{5}.\end{aligned} \tag{2.3.6}$$

确定这个相同的后验概率的第二种方法是在给定 H_1 的条件下，利用贝叶斯定理(2.3.4)的条件形式. 在 H_1 的条件下，B_1 的条件概率为1/3，从而 B_2 的条件概率是2/3. 这些条件概率现在可以作为下一阶段(即第二次投掷硬币时)的先验概率. 从而，在(2.3.4)中令 $C = H_1$，$\Pr(B_1 \mid H_1) = 1/3$，$\Pr(B_2 \mid H_1) = 2/3$，然后我们计算出后验概率 $\Pr(B_1 \mid H_1H_2)$. 这是在第二次投掷时观测到一个头像和第一次投掷时观测到一个头像之后硬币是均匀硬币的概率. $\Pr(H_2 \mid B_1H_1)$ 是我们需要的. 既然 H_1 与 H_2 在 B_1 条件下是独立的，那么依据定理2.2.3，$\Pr(H_2 \mid B_1H_1) = \Pr(H_2 \mid B_1) = 1/2$. 既然当 B_2 发生时，硬币是有两个头像的，那么 $\Pr(H_2 \mid B_2H_1) = 1$，从而我们得到

$$\begin{aligned}\Pr(B_1 \mid H_1H_2) &= \frac{\Pr(B_1 \mid H_1)\Pr(H_2 \mid B_1H_1)}{\Pr(B_1 \mid H_1)\Pr(H_2 \mid B_1H_1) + \Pr(B_2 \mid H_1)\Pr(H_2 \mid B_2H_1)} \\ &= \frac{(1/3)(1/2)}{(1/3)(1/2) + (2/3)(1)} = \frac{1}{5}.\end{aligned} \tag{2.3.7}$$

用第二种方法得到的事件 B_1 的后验概率跟第一种方法得到的一样. 我们有下面一般性的结论：如果一个试验多阶段进行，那么在多阶段中也能计算每个事件的后验概率. 在每个阶段进行之后，这个阶段的后验概率可以作为下一个阶段的先验概率. 读者可以回看(2.3.4)，即为贝叶斯定理的条件形式. 投掷硬币的例子就是贝叶斯定理和它的条件形式诸多应用的一个典型，因为我们假设在分割的每个元素 $B_1,\cdots,B_k$（在上面例子中，$k=2$）条件下观测事件都是独立的. 而条件独立性使得在 B_1 或是 B_2 条件下 H_i（第 i 次投掷有头像）的概率在考虑或是不考虑前期投掷时都是一样的（参见定理2.2.3）.

2.3.5 扩展研究：条件独立事件

计算导出的(2.3.5)，(2.3.7)与例2.2.9一起说明了观测事件的一个有用的统计模型的简单情况. 通常会遇到一列类似的事件，它们有相同的发生概率. 通常还会有事件的标记次序不影响我们给定的概率. 然而，我们通常认为这些事件不是独立的，因为依据早期事件发生的个数我们就会改变后来发生事件的概率. 例如，在本节前面投掷硬币的计算中，在任意投掷发生之前，H_2 的概率和 H_1 的概率是一样的，根据定理2.3.1，(2.3.5)的分母是3/4. 然而，在观测到事件 H_1 发生之后，H_2 的概率为 $\Pr(H_2|H_1)$，由(2.3.1)，(2.3.7)的分母是5/6. 即使我们把在硬币为均匀硬币条件下投掷硬币视为条件独立的，在硬币为两个头像时也视为条件独立的（在这种情况下我们总能知道每次发生的事情），但是如果没有条件信息，我们也不能视它们为独立的. 因为条件信息把问题中不确定性的主要来源去掉了，所以我们可以相应地来划分样本空间. 现在，在分割事件的条件下，利用投掷的条件独立性来计算事件不同组合的联合概率. 最后，用定理2.3.1和(2.3.1)把这些概率联系起来. 多个例子可以说明这一思想.

73

例2.3.9 了解概率 在例2.2.9中，机器以 $p=0.01$ 或 $p=0.04$ 的概率生成次品部件，假设 $p=0.01$ 的先验概率是0.9. 在随机抽取六个部件后，假设我们看到两个次品. 问 $p=0.01$ 的后验概率是多少？

在例2.2.9中，令 $B_1=\{p=0.01\}$，$B_2=\{p=0.4\}$，A 是大小为6的随机样本中有两个次品的事件. B_1 的先验概率是0.9，B_2 的先验概率是0.1. 在例2.2.9中，我们已经计算出 $\Pr(A|B_1)=1.44\times10^{-3}$和 $\Pr(A|B_2)=0.311$，由贝叶斯定理得：

$$\Pr(B_1|A)=\frac{0.9\times1.44\times10^{-3}}{0.9\times1.44\times10^{-3}+0.1\times0.311}=0.04.$$

即使我们最初认为 B_1 有高达0.9的概率，在我们得知在一个小至6的样本中有两个次品之后，我们明显改变想法，现在我们认为 B_1 有小至0.04的概率. 改变这么大的原因是如果 B_1 是真的而非 B_2 是真的，那么事件 A 发生的可能性更大. ——◀

例2.3.10　诊疗试验　考虑例2.3.2和例2.3.4中描述的相同的诊疗试验，令 E_i 是第 i 个病人结果为成功的事件. B_j 仍是概率为 $p=(j-1)/10$（对 $j=1,\cdots,11$ 都成立）的事件，这里 p 是成功的概率. 如果知道 p，则称 $E_1,E_2,\cdots$ 是相互独立的，即，我们要模拟病人关于 p 的每个值条件独立. 令 $\Pr(E_i|B_j)=(j-1)/10$ 对所有 i，j 成立. 我们仍然假设在试验开始之前对所有 j 成立 $\Pr(B_j)=1/11$. 现在，我们可以通过计算每个病人完成试验之后 B_j 事件的后验概率来表示我们了解 p 的程度.

74

例如，考虑第一个病人，在(2.3.2)中计算出 $\Pr(E_1)=1/2$，如果 E_1 发生，应用贝叶斯定理，我们得到

$$\Pr(B_j|E_1)=\frac{\Pr(E_1|B_j)\Pr(B_j)}{1/2}=\frac{2(j-1)}{10\times 11}=\frac{j-1}{55}. \tag{2.3.8}$$

在观测到一个病人是成功的之后，如我们所预料的，大的 p 值的后验概率大于它们的先验概率，小的 p 值的后验概率小于它们的先验概率. 例如，$\Pr(B_1|E_1)=0$，因为在一个人成功之后，$p=0$ 被排除掉了. $\Pr(B_2|E_1)=0.0182$，而比它先前的值0.0909小得多，$\Pr(B_{11}|E_1)=0.1818$，而比它先前的值0.0909大得多.

我们能够检查在每个病人被观测后后验概率是如何变化的. 然而，我们应该跳到表2.1的栏所有40个病人已经被观测的点前面. A 代表他们中的22个是成功的，18个是失败的观测事件，我们利用跟例2.2.4相同的方法来计算 $\Pr(A|B_j)$. 40个病人有22个成功者，形成 $\binom{40}{22}$ 个可能的序列，并且在 B_j 的条件下，每一个序列的概率是 $([j-1]/10)^{22}(1-[j-1]/10)^{18}$. 因此，对每个 j 有

$$\Pr(A|B_j)=\binom{40}{22}([j-1]/10)^{22}(1-[j-1]/10)^{18},$$

然后，由贝叶斯定理，

$$\Pr(B_j|A)=\frac{\frac{1}{11}\binom{40}{22}([j-1]/10)^{22}(1-[j-1]/10)^{18}}{\sum_{i=1}^{11}\frac{1}{11}\binom{40}{22}([i-1]/10)^{22}(1-[i-1]/10)^{18}}.$$

图2-3表明在观测到 A 后11个分割元素的后验概率. 注意到 B_6 和 B_7 的概率最大，为0.42. 这与观测样本中成功的比率 $22/40=0.55$ 是一致的，介于 $(6-1)/10$ 和 $(7-1)/10$ 之间.

75

我们也能计算在试验之前且在40个病人之后，下一个病人是成功的概率. 试验之前，按(2.3.2)计算的结果，$\Pr(E_{41})=\Pr(E_1)=1/2$. 在观测40个病人之后，通过使用全概率公式的条件形式(2.3.1)，可以计算 $\Pr(E_{41}|A)$ 为：

$$\Pr(E_{41}|A)=\sum_{j=1}^{11}\Pr(E_{41}|B_jA)\Pr(B_j|A). \tag{2.3.9}$$

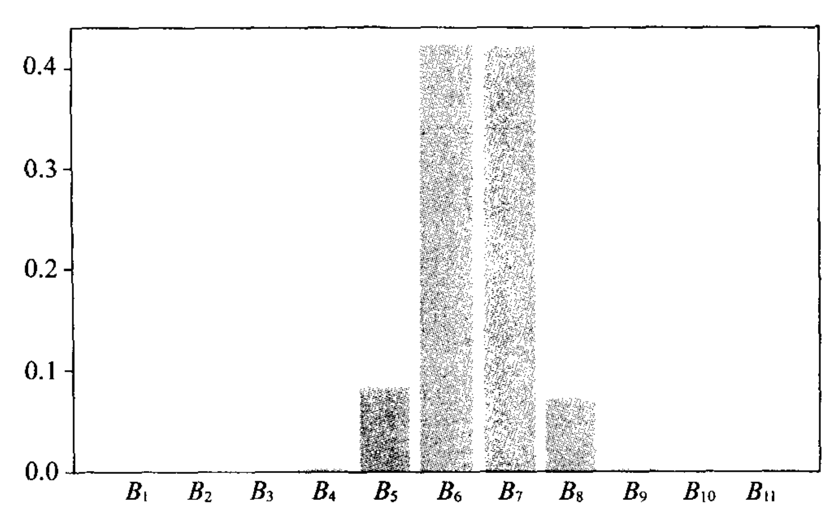


图 2-3　例 2.3.10 中 40 个病人之后分割元素的后验概率

利用图 2-3 中 $\Pr(B_j|A)$ 的数据，以及 $\Pr(E_{41}|B_jA)=\Pr(E_{41}|B_j)=(j-1)/10$（$E_i$ 关于 B_j 的条件独立性），代入(2.3.9)计算得 0.5476. 这与观测的成功频率是非常接近的. ◀

例 2.3.2 中最后的计算结果是在观测到很多有着相同条件概率的条件独立事件之后发生了什么. 下一个事件关于观测事件的条件概率趋近于在观测事件中事件发生的频率. 的确，当有实际数据时，先验概率的选择变得一点都不重要.

例 2.3.11　先验概率的影响　考虑与例 2.3.10 中相同的临床试验，假设不同的调查者关于成功的概率值有不同的先验看法. 该调查者认为有下面的先验概率.

事件	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}
p	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
先验概率	0.00	0.39	0.20	0.13	0.09	0.07	0.05	0.03	0.02	0.02	0.00

我们重新用贝叶斯定理计算后验概率，得到图 2-4 中的数据. 为了有助于比较，例 2.3.10 中的后验概率也在图 2-4 中用符号×标出. 可以看出不管在先验概率之间有多么不同，后验概率的两个集合却那么接近. 如果过去有很少病人被观测，那么在后验概率的两个集合之间会有较大的不同，因为被观测到的事件提供较少的信息（见习题 16）. ◀

76

2.3.6　小结

分割是互不相交的事件的集合，这些互不相交的事件的并集构成整个样本空间. 要使分割有用，选择的分割应该满足：如果我们知道分割事件中的一个发生，那么就能去掉不确定性的一个重要来源. 如果事件 A 关于分割中的每个事件的条件概率

是有意义的，全概率公式告诉我们怎样去联合这些条件概率以得到 $\Pr(A)$. 贝叶斯定理告诉我们怎样去计算每个分割事件关于 A 的条件概率. 分割的一个重要应用是把样本空间分割成足够小的部分，以至于一系列有用事件关于分割中的每个事件都是条件独立的.

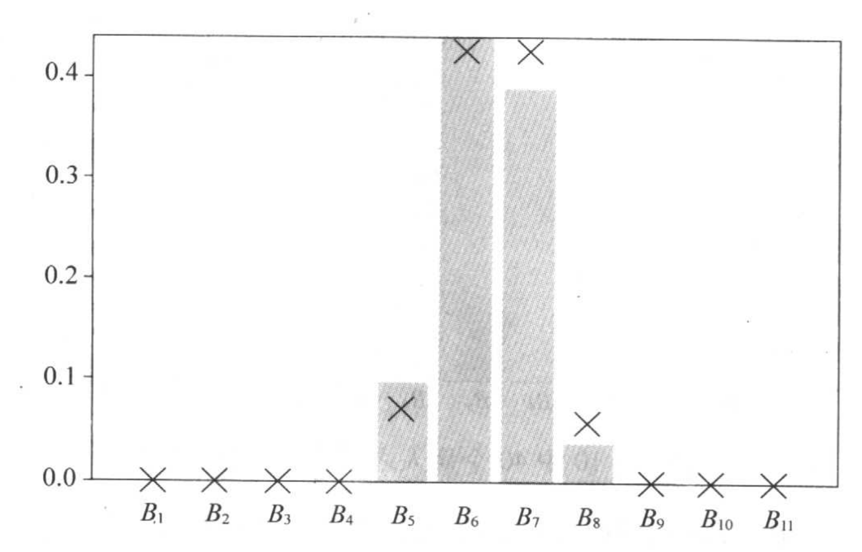


图 2-4　例 2.3.11 中 25 个病人之后分割元素的后验概率，×表示例 2.3.10 中计算的后验概率的值

习题

1. 盒子里有三个每面都有头像的硬币，四个每面都是文面的硬币，还有两个均匀硬币. 如果随机选取九个硬币中的一个，并且投掷一次，问得到一个头像的概率是多少？
2. 由于修理状况不同机器生产次品部件服从三种不同的概率. 如果机器正常运作，它以概率 0.02 生产次品部件. 如果机器老化，它以概率 0.1 生产次品部件. 如果它需要维修，它以概率 0.3 生产次品部件. 机器正常运作的概率为 0.8，老化的概率为 0.1，需要维修的概率为 0.1. 求随机抽取一个部件是次品的概率.
3. 在三个不同的选举区自由党投票人的百分比进行如下划分：一区 21%，二区 45%，三区 75%. 如果随机抽取一个区，从这个区中随机抽取一个投票人，问他是自由党的概率是多少？
4. 重新考虑 2.1 节中习题 4 中的购物者. 每一次购物，他选择和他先前购买的同样商标的牙膏的概率是 1/3，他选择改变商标的概率为 2/3. 假定在第一次购买中，他选择商标 A 的概率为 1/4，选择商标 B 的概率为 3/4. 问他的第二次选择是商标 B 的概率为多少？
5. 假定 k 个事件 $B_1,\cdots,B_k$ 构成样本空间 S 的一个分割. 对每个 $i=1,\cdots,k$，用 $\Pr(B_i)$ 表示 B_i 的先验概率. 对每个事件 A 满足 $\Pr(A)>0$，用 $\Pr(B_i|A)$ 表示 B_i 关于事件 A 的后验概率. 证明：如果 $\Pr(B_1|A)<\Pr(B_1)$，那么至少有一个 $i(i=2,\cdots,k)$ 满足 $\Pr(B_i|A)>\Pr(B_i)$.
6. 再次考虑例 2.3.7 中的条件. 从一批生产部件中随机抽取一个部件，发现是次品. 问：i 取何值时 $(i=1,2,3)$ 满足：由机器 M_i 生产的部件的后验概率大于由机器 M_i 生产的部件

的先验概率?

7. 假定在本节例 2.3.7 中，从整个部件中随机选取的一个部件是合格品. 确定是由机器 M_2 生产的后验概率.

8. 设计一项新的试验来检查一种特殊类型的癌症. 如果该项测试用于一个得这种典型癌症的人，这个人有阳性反应的概率为 0.95，有阴性反应的概率为 0.05. 如果该项测试用于一个没有这种典型癌症的人，这个人有阳性反应的概率为 0.05，有阴性反应的概率为 0.95. 假定在一般人群中，每 100 000 个人中有 1 个人得这种典型的癌症，如果随机选取一个人，这个人对测试有阳性反应，那么问他得这种典型癌症的概率是多少?

9. 某一个城市，30% 的人是保守党，50% 是自由党，20% 是无党派人士. 有记录显示在某一次选举中，65% 的保守党投票，82% 的自由党投票，50% 的无党派人士投票. 如果在城市中随机选取一个人，他在上次选举中没有投票，问：他是自由党的概率是多少?

10. 假定适当调整一个机器时，它生产部件的 50% 是高质量的，另外 50% 是中间质量的. 然而，假定这个机器不恰当调整，在这种情况下，由它生产的部件的 25% 是高质量的，75% 是中间质量的.
    a. 假定在某一时间随机选取由机器生产的 5 个部件进行检测. 如果有 4 个是高质量的，1 个是中间质量的，问这时机器得到适当调整的概率是多少?
    b. 假设再选一个与其他 5 个同时由机器生产的部件，发现它是中间质量. 问：机器得到适当调整的新的后验概率是多少?

11. 假定一个盒子有 5 个硬币，对每个硬币投掷得到头像有着不同的概率. 令 p_i 表示当投掷第 $i(i=1,\cdots,5)$ 个硬币时得到头像的概率. 假定 $p_1=0$，$p_2=1/4$，$p_3=1/2$，$p_4=3/4$，$p_5=1$.
    a. 假定从盒子里随机选取一个硬币，投掷一次得到一个头像. 问：选取第 $i(i=1,\cdots,5)$ 个硬币的后验概率是多少?
    b. 如果再次投掷这个硬币，问得到另一个头像的概率是多少?
    c. 选取一个硬币进行第一次投掷得到一个文面，同样的硬币再次投掷，问在第二次投掷时得到头像的概率是多少?

12. 再次考虑习题 11 中的一个有着 5 个不同硬币的盒子. 假定从盒子中随机选取一个硬币，进行重复投掷，直至出现头像.
    a. 如果在第四次投掷时得到第一个头像，问选取第 $i(i=1,\cdots,5)$ 个硬币的后验概率是多少?
    b. 如果继续投掷这个硬币直到得到另一个头像，问正好需要再投掷三次的概率是多少?

13. 证明全概率公式(2.3.1)的条件形式.

14. 再次考虑例 2.3.8 的情况. 检测个体的表型，并且呈现显性特征. 求 $i(i=1,\cdots,6)$值，使其满足：父母为事件 B_i 的基因类型的后验概率比父母为事件 B_i 的基因类型的先验概率小?

15. 在例 2.3.8 中假定检测个体呈隐性特征，试确定父母有事件 B_4 基因类型的后验概率.

16. 在例 2.3.10 和例 2.3.11 的诊疗试验中，假定我们仅仅观测到前 5 个病人，并且有 3 个是成功的. 利用例 2.3.10 和例 2.3.11 中两组不同的先验概率来计算两组后验概率. 问：这两组后验概率是否与例 2.3.10 和例 2.3.11 中的两组一样彼此接近? 为什么?

17. 假定一个盒子有一个均匀硬币和一个每面都有一个头像的硬币，从盒子里随机抽取一个

硬币进行投掷. 在方程式(2.3.6)和(2.3.7)中，计算在前两次投掷都产生头像的条件下硬币是均匀的条件概率.

a. 假定第三次投掷硬币又得到一个头像，计算在所有三次投掷都产生头像的条件下硬币是均匀的概率；

78

b. 假定第四次投掷硬币，结果是文面，计算硬币是均匀的后验概率.

18. 再次考虑2.2节习题23中的情况. A 是11个程序中正好有8个得到编译的事件. 计算在条件 B 下 A 的条件概率.
19. 利用例2.3.11中对事件 $B_1,\cdots,B_{11}$ 的先验概率. 令 E_1 是第一个病人是成功的事件，计算 E_1 的概率并且说明为什么它比例2.3.10中计算的数值小得多.
20. 再次考虑习题2中的情况. 假定观测到几个部件，不同部件在机器修理的三个状态下都是条件独立的，计算这三个修理状态的后验概率.

2.4　马尔可夫链

随时间以随机方式变化的系统的一个常用模型是马尔可夫链模型. 在马尔可夫链中，给定过去状态和现在状态条件下的将来状态的条件概率仅依赖于现在的状态.

2.4.1　随机过程

假设某一商务办公室有五条电话线，在任意指定时间可以使用这些电话线中的任意多条. 在规定的2分钟时间内，观察电话线路，记录下每次使用的电话线数目. 用 X_1 表示在这个时段开始时，第一次观察到的使用的电话线数目，用 X_2 表示第二次观察到的使用的电话线数目，一般地，当 $n=1,2,\cdots$，用 X_n 表示第 n 次观察到的使用的电话线数目.

称观察到的这个序列 $X_1,X_2,\cdots$，为一个*随机过程*. 由于不能精确地预知这些观察到的值，但是能确定在任意特定时间内它们每一个不同的可能值的概率. 称刚才所描述的随机过程为*离散时间参数随机过程*. 因为观察使用电话线的时刻只是在离散的或分散的时间点上，而不是连续时间.

在随机过程中，称第一次观察的 X_1 为过程的*初始状态*；对于 $n=2,3,\cdots$，称观察到的 X_n 为 *n 时刻过程的状态*. 在前述的例子中，任意时刻状态是在那个时刻使用电话线的数目，因此每个状态必定是0和5之间的一个整数. 在本章余下的部分中，我们将考虑的是在任意指定时刻可能状态只是有限数目的随机过程.

在离散时间参数随机过程中，过程的状态从一个时刻到另一个时刻随机地变化. 为了描述一个特定过程的完备的概率模型，需要确定在初始状态 X_1 的每一个可能取值的概率. 也要确定对每个后续状态 $X_{n+1}(n=1,2,\cdots)$ 具有以下形式的每个条件概率：

$$\Pr(X_{n+1}=x_{n+1}\mid X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$$

换句话说，对每个时刻 n，概率模型必须是能确定出在给定时刻 $1,\cdots,n$ 时，过

程状态是 $x_1;\cdots,x_n$，在时刻 $n+1$ 时的状态是 x_{n+1} 的条件概率. [79]

2.4.2 马尔可夫链

定义：一个马尔可夫链是随机过程序列 $X_1,X_2,\cdots$，的一种特殊类型，可以描述如下：在任意时刻 n，当现在状态 X_n 和所有以前 $X_1,\cdots,X_{n-1}$ 的过程是已知的，所有将来的状态 $X_j(j>n)$ 的概率仅依赖于现在状态 X_n，而并不依赖以前的状态 $X_1,\cdots,X_{n-1}$. 形式上，一个马尔可夫链是这样一个随机过程，它对于 $n=1,2,\cdots$ 的每个可能状态序列 $x_1,x_2,\cdots,x_{n+1}$，有

$$\Pr(X_{n+1}=x_{n+1}\mid X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$$
$$=\Pr(X_{n+1}=x_{n+1}\mid X_n=x_n).$$

由 2.1 节中给出的条件概率的乘法规则，马尔可夫链中的概率必须满足下面的关系

$$\Pr(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$$
$$=\Pr(X_1=x_1)\Pr(X_2=x_2\mid X_1=x_1)\Pr(X_3=x_3\mid X_2=x_2)\cdots\Pr(X_n=x_n\mid X_{n-1}=x_{n-1}).$$

例 2.4.1 购买牙膏 在 2.1 节的习题 4 中，我们考虑一位顾客在几种场合下选择两种品牌的牙膏，如果顾客在第 i 次购买中选择品牌 A，就记作 $X_i=1$，如果顾客在第 i 次购买中选择品牌 B，就记作 $X_i=2$. 那么状态的序列 $X_1,X_2,\cdots$ 是一个每次都有两种可能状态的随机过程. 要确定购买的概率，设顾客本次与前一次购买中选择相同品牌的概率是 1/3，与前一次购买选择不同品牌的概率是 2/3. 因为本次与前一次以前的各次无关，我们就看到这个随机过程是一个有如下条件概率的马尔可夫链

$$\Pr(X_{n+1}=1\mid X_n=1)=1/3,\Pr(X_{n+1}=2\mid X_n=1)=2/3,$$
$$\Pr(X_{n+1}=1\mid X_n=2)=2/3,\Pr(X_{n+1}=2\mid X_n=2)=1/3.$$ ◀

有平稳转移概率的马尔可夫链

我们现在考虑一个只取有限数目 k 个可能状态 $s_1,\cdots,s_k$ 的马尔可夫链，在每个时刻这个链必须处于这 k 个状态之一. 称这种类型的一个马尔可夫链为一个有限马尔可夫链.

称马尔可夫链在 n 时刻处于状态 s_i 条件下，在 $n+1$ 时刻的状态是 s_j 的条件概率 $\Pr(X_{n+1}=s_j\mid X_n=s_i)$ 为转移概率. 如果确定马尔可夫链的转移概率在每一时刻 $n(n=1,2,\cdots)$ 都有相同的值. 那么就说这个马尔可夫链有平稳转移概率. 换句话说，一个马尔可夫链有平稳转移概率，如果对所有状态 s_i 和状态 s_j，存在状态转移概率 p_{ij} 满足

$$\Pr(X_{n+1}=s_j\mid X_n=s_i)=p_{ij}\quad 对于\ n=1,2,\cdots$$

那么例 2.4.1 中的马尔可夫链有平稳转移概率. [80]

为了举例说明这些定义的应用，我们再次考虑含有五条电话线的办公室，为了使这个随机过程是个马尔可夫链，在每时刻每个可能在使用的电话线数目的概率必

须只依赖于在最近观察的2分钟的过程中所使用的线路数目，而不依赖于任何以前观察的结果. 例如，在时刻 n，如果有三条线在使用，不管在 $n-1$ 时刻在使用的线路是0、1、2、3、4或5，在 $n+1$ 时刻的概率必定是相同的. 不过，在现实中 $n-1$ 时刻观察的结果可以提供关于在 n 时刻使用的三条线中每一条被占用的时间长度的一些信息，这些信息能有助于确定在 $n+1$ 时刻的概率. 然而，我们可假设现在的这个过程是马尔可夫链. 如果这个马尔可夫链有平稳转移概率，打进与打出电话的使用率必定是确定的，包括在这个过程中的整个期间，这些电话呼叫的平均持续时间不会改变. 这个要求意味着在整个期间：不能包括有更多的呼叫繁忙时刻的要求或很少有呼叫的空闲时段. 例如，如果在一个特定的时刻只有一条电话线在使用，不管这是发生在整个过程的哪个时期，在2分钟后恰有 j 条电话线被使用的概率必定是 p_{1j}.

2.4.3　转移概率

1. 一步转移概率

考虑一个具有 k 种可能状态 $s_1,\cdots,s_k$ 和平稳转移概率的马尔可夫链. 对于 $i=1,\cdots,k$ 和 $j=1,\cdots,k$，我们再次用 p_{ij} 表示如果它在以前观察时刻的状态为 s_i，在指定观察时刻的状态为 s_j 的条件下的条件概率. 马尔可夫科的**转移矩阵**被定义为元素是 p_{ij} 的 $k\times k$ 阶矩阵 P. 有：

$$P=\begin{bmatrix} p_{11} & \cdots & p_{1k} \\ p_{21} & \cdots & p_{2k} \\ \vdots & \cdots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}. \tag{2.4.1}$$

因为每个数 $p_{ij}\geqslant 0$，而且 $\sum_{j=1}^{n} p_{ij}=1$，对 $i=1,\cdots,k$ 成立. 因为在给定的观察时刻，该链的状态是 s_i，那么在下一个观察时刻每个状态 $s_1,\cdots,s_k$ 的概率之和必定是1.

称所有元素是非负且每行和为1的一个方阵为**随机矩阵**. 可见具有平稳转移概率的任何确定马尔可夫链的转移矩阵 P 必定是一个随机矩阵. 反之，任意 $k\times k$ 随机矩阵能用作具有 k 种可能状态和平稳转移概率的一个马尔可夫链的转移矩阵. 对于例2.4.1中的马尔可夫链的转移矩阵是

$$P=\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

81

例2.4.2　使用电话线数目的转移矩阵　假设在有五条电话线的办公室的例子中，在时刻 $n=1,2,\cdots$，正在使用的电话线数目构成一个具有平稳转移概率的马尔可夫链形式. 这个链有六种可能的状态 $b_0,b_1,\cdots,b_5$，这里的 b_i 是在给定的时刻恰有 i 条线在使用的状态($i=0,1,\cdots,5$). 假设转移矩阵 P 有如下形式：

$$P = \begin{array}{c} \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{array} \begin{array}{c} \begin{array}{cccccc} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \end{array} \\ \left[\begin{array}{cccccc} 0.1 & 0.4 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.4 & 0.2 \end{array}\right] \end{array}. \tag{2.4.2}$$

(a) 假设在一个确定的观察时刻所有五条线都在使用，我们就能确定在下一个观察时刻恰好有四条线在使用的概率.

(b) 假设在一个确定的时刻一条线都没使用，我们就能确定在下一个观察时刻至少有一条线在使用的概率.

(i) 这个概率是矩阵 P 中对应于状态 b_5 的行和对应于状态 b_4 的列的元素，它的值是0.4.

(ii) 如果在一个确定的时刻没有电话线路在使用，那么矩阵 P 的左上角的元素就是在下一个观察时刻没有线路在使用的概率. 它的值是0.1. 那么在下一个观察时刻至少有一条线在使用的概率是 $1-0.1=0.9$. ◀

例2.4.3 植物繁殖实验 植物学家正在研究一种雌雄同株的植物的变化(在单个植物花朵上既有雄性器官又有雌性器官). 它开始时有两种植物Ⅰ和Ⅱ相互传授花粉，雄性的Ⅰ传授花粉给雌性的Ⅱ，雌性的Ⅰ传授花粉给雄性的Ⅱ. 这样就产生了下一代的两个个体. 原来的植物被破坏了，等到新一代的两个个体一成熟，以上过程又重新开始. 研究的几个复制同时进行，植物学家可能感兴趣的是，在每一个子代中，对于每个特定基因的几种可能的基因型植物的比例. 假设一个基因型有两个等位基因 A 和 a，一个个体的基因型是以下三种类型AA，Aa或aa中的一种. 当新的个体出生时，它从父代中的一个得到两个等位基因中的一个(得到每个基因的概率是1/2)，从父代另外一个个体独立地得到两个等位基因中的一个. 两个后代分别独立地得到他们各自的基因型. 例如，如果这个父代有基因型AA和Aa，那么一个后代就肯定从父代中第一个母体得到A，从第二个母体分别以1/2的概率得到A或a. 设这个种群的状态为现在种群中的两个成员的基因类型的集合，对集合{AA,Aa}和{Aa,AA}不加区分，这样就有六种状态：{AA,AA}，{AA,Aa}，{AA,aa}，{Aa,Aa}，{Aa,aa}和{aa,aa}. 我们能够计算下一代是六种状态中的哪一种状态的概率. 例如，如果状态是{AA,AA}或{aa,aa}中的一种，则下一代将是相同状态的概率是1. 如果状态是{AA,aa}，则下一代是状态{Aa,Aa}的概率也是1，其他三种状态有更为复杂的转移概率. 82

如果现在的状态是{Aa,Aa}，那么下一代就可能是六种状态中的任意一种. 为了计算六种转移概率，首先要计算一个给定的后代会有三种基因型任一种的概率，

图2-5指明了这种状态可能的后代．在图2-5中每一个向下箭头就是一个等位基因的可能遗传，箭头的每个组合终止于一个基因型的概率是1/4. 由此可知AA和aa的概率都是1/4，而Aa的概率是1/2.

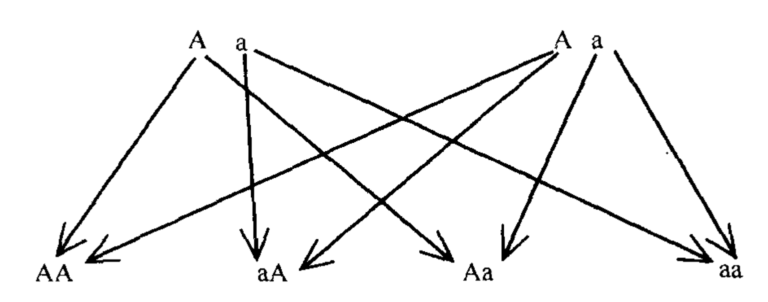


图2-5　按{Aa,Aa}的繁殖

因为有两个箭头的不同组合产生了该后代．为了使下一个状态是{AA,AA}，这两个后代必定分别独立地是AA，所以这个转移概率是1/16，同理可得转移到{aa,aa}的概率也是1/16，到{AA,Aa}的一个转移要求一个后代是AA(概率1/4)，另一个是Aa(概率1/2)，但这两个不同的遗传类型可以任何顺序出现，所以该转移总的概率是$2\times(1/4)\times(1/2)=1/4$. 同理可得，转移到{Aa,aa}的概率也是1/4. 转移到{AA,aa}需要一个后代是AA(概率1/4)和另一个后代是aa(概率1/4). 另外，这些过程可以以两种顺序发生．所以总的概率是$2\times1/4\times1/4=1/8$. 利用减法转移到{Aa,Aa}的概率必定是$1-1/16-1/16-1/4-1/4-1/8=1/4$. 这是一个完整转移矩阵，可以用上述类似的方法来验证它.

	{AA, AA}	{AA,Aa}	{AA,aa}	{Aa,Aa}	{Aa, aa}	{aa,aa}
{AA, AA}	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
{AA, Aa}	0.2500	0.5000	0.0000	0.2500	0.0000	0.0000
{AA, aa}	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
{Aa, Aa}	0.0625	0.2500	0.1250	0.2500	0.2500	0.0625
{Aa, aa}	0.0000	0.0000	0.0000	0.2500	0.5000	0.2500
{aa, aa}	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

◀

2. 多步转移矩阵

83

再次考虑k种可能的状态$s_1,\cdots,s_k$的任意马尔可夫链和由等式(2.4.1)所得的转移矩阵P，并且假设这个链是在时刻n的状态处于s_i中，我们来确定链在时刻$n+2$处于状态s_j的概率．换句话说，我们来确定从状态s_i到s_j的两步转移概率，记这个概率为$p_{ij}^{(2)}$.

对于$n=1,2,\cdots$，设X_n表示在n时刻链的状态，如果s_r表示链在$n+1$时刻转移到的状态，那么

$$
\begin{aligned}
p_{ij}^{(2)} &= \Pr(X_{n+2}=s_j \mid X_n=s_i)\\
&= \sum_{r=1}^{k} \Pr(X_{n+1}=s_r, X_{n+2}=s_j \mid X_n=s_i)\\
&= \sum_{r=1}^{k} \Pr(X_{n+1}=s_r \mid X_n=s_i)\Pr(X_{n+2}=s_j \mid X_{n+1}=s_r, X_n=s_i)\\
&= \sum_{r=1}^{k} \Pr(X_{n+1}=s_r \mid X_n=s_i)\Pr(X_{n+2}=s_j \mid X_{n+1}=s_r)\\
&= \sum_{r=1}^{k} p_{ir}p_{rj},
\end{aligned}
$$

这里第三个等式可以由定理2.1.2得到，第四个等式可以由马尔可夫链的定义得到.

$p_{ij}^{(2)}$的值可由以下方式决定：如果将转移矩阵P取平方，即构造矩阵$P^2=PP$. 那么矩阵P^2的第i行第j列的元素将是$\sum_{r=1}^{k} p_{ir}p_{rj}$. 因此$p_{ij}^{(2)}$将是$P^2$的第$i$行第$j$列的元素.

用类似的推理，这个链从状态s_i经过以三步转移到s_j的概率就是$p_{ij}^{(3)}=\Pr(X_{n+3}=s_j \mid X_n=s_i)$，可以通过构造矩阵$P^3=P^2P$来得到. 那么概率$p_{ij}^{(3)}$即为矩阵$P^3$的第$i$行第$j$列的元素.

一般地，对于$m(m=2,3,\cdots)$的每一个值，矩阵P的m次幂P^m确定了这个链从状态s_i通过m步转移到状态s_j的概率$p_{ij}^{(m)}$，因此称矩阵P^m为马尔可夫链的m步转移矩阵.

例2.4.4 使用电话线数目的两步和三步转移矩阵 再次考虑由等式(2.4.2)给出的占用五条电话线的马尔可夫链的转移矩阵. 我们首先假设在某一特定时刻有i条线在使用，我们将确定两个时间段后恰有j条线在使用的概率. 如果我们将矩阵P乘以自己，就得到以下的两步转移矩阵：

84

$$
P^2 = \begin{array}{c} \\ b_0\\ b_1\\ b_2\\ b_3\\ b_4\\ b_5 \end{array}
\begin{array}{c}
\begin{array}{cccccc} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \end{array}\\
\begin{bmatrix}
0.14 & 0.23 & 0.20 & 0.15 & 0.16 & 0.12\\
0.13 & 0.24 & 0.20 & 0.15 & 0.16 & 0.12\\
0.12 & 0.20 & 0.21 & 0.18 & 0.17 & 0.12\\
0.11 & 0.17 & 0.19 & 0.20 & 0.20 & 0.13\\
0.11 & 0.16 & 0.16 & 0.18 & 0.24 & 0.15\\
0.11 & 0.16 & 0.15 & 0.17 & 0.25 & 0.16
\end{bmatrix}
\end{array}. \qquad (2.4.3)
$$

从这个矩阵我们可以得到该链的如下的任意两步转移概率：

i. 如果在某一指定时刻使用两条电话线，那么在两个时段后有四条电话线在使用的概率是0.17.

ii. 如果在某一指定时刻使用三条电话线，那么在两个时段后有三条电话线在使用的概率是0.20.

现在假设在某一确定时间使用 i 条电话线，我们来确定在以后的三个时段后恰好使用 j 条电话线的概率．如果我们构造矩阵 $P^3=P^2P$，就得到以下三步转移矩阵：

$$P^3=\begin{array}{c} \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{array}\begin{array}{c} \begin{array}{cccccc} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \end{array} \\ \begin{bmatrix} 0.123 & 0.208 & 0.192 & 0.166 & 0.183 & 0.128 \\ 0.124 & 0.207 & 0.192 & 0.166 & 0.183 & 0.128 \\ 0.120 & 0.197 & 0.192 & 0.174 & 0.188 & 0.129 \\ 0.117 & 0.186 & 0.186 & 0.179 & 0.199 & 0.133 \\ 0.116 & 0.181 & 0.177 & 0.176 & 0.211 & 0.139 \\ 0.116 & 0.180 & 0.174 & 0.174 & 0.215 & 0.141 \end{bmatrix} \end{array}. \tag{2.4.4}$$

从这个矩阵可以得到该链的任意三步转移概率，例如：

i. 如果在某一指定时刻所有五条电话线都在使用，那么在三个时段后没有使用线路的概率是0.116.

ii. 如果在某一指定时刻使用一条电话线，那么在三个时段后恰好使用一条线的概率是0.207. ◀

85

例2.4.5 植物繁殖实验 在例2.4.3中，转移矩阵有许多0元素，因为许多转移没有发生．然而，如果我们愿意等两步，将会发现在两步中不可能出现的转移只有从第一个状态到任意其他状态，或者从最后一个状态到任意其他状态的转移．确实，如果我们观察三步、四步、或 m 步的转移矩阵，第一行和最后一行总是相同的，这两种状态就叫做吸收态．因为马尔可夫链一旦进入这些状态就不能脱离这种状态．既然不管链从哪开始，以有限步进入吸收态的概率是正的．如果允许链运行足够长时间，则它被最终吸收到吸收态的概率就是1. ◀

2.4.4 初始概率向量

假设一个有平稳转移概率的有限马尔可夫链，有 k 种可能状态 $s_1,\cdots,s_k$，在初始观察时刻 $n=1$ 时，这个链的状态可以是这 k 种状态中的任意一种．假设对于 $i=1,\cdots,k$，在该过程开始时，该链处于状态 s_i 的概率是 v_i，这里 $v_i\geqslant 0$ 且 $v_1+\cdots+v_k=1$.

一个向量 $\boldsymbol{w}=(w_1,\cdots,w_n)'$，如满足对 $i=1,\cdots,k$，$w_i\geqslant 0$，且 $\sum_{i=1}^k w_i=1$，则称它为一个概率向量．这个概率向量 $\boldsymbol{v}=(v_1,\cdots,v_k)$，给出了在初始观察时刻链的各个状态的概率，称它为这个链的*初始概率向量*．对例2.4.1，在2.1节的习题4给出了它的初始概率向量为 $\boldsymbol{v}=(0.5,0.5)$.

初始概率向量和转移矩阵共同决定了这个链在每个特定时刻处于每个特定状态的概率．如果 $\boldsymbol{v}$ 是链的初始概率向量，那么 $\Pr(X_1=s_i)=v_i,(i=1,\cdots,k)$．如果这个链的转移矩阵是 $k\times k$ 阶矩阵 P，它的元素 p_{ij} 由等式(2.4.1)来表示．那么对于 $j=1,\cdots,k$，有

$$\Pr(X_2=s_j)=\sum_{i=1}^k\Pr(X_1=s_i,X_2=s_j)=\sum_{i=1}^k\Pr(X_1=s_i)\Pr(X_2=s_j\mid X_1=s_i)=\sum_{i=1}^k v_ip_{ij}.$$

因为 $\sum_{i=1}^{k} v_i p_{ij}$ 是向量 $\boldsymbol{v}P$ 的第 j 个分量，这个推导表明在观察时刻 2，该链的这种状态概率是由概率向量 $\boldsymbol{v}P$ 所确定的.

更一般地，假设在某个给定时刻 n，该链处于状态 s_i 的概率是 $\Pr(X_n = s_i) = w_i$，$i=1,\cdots,k$. 则

$$\Pr(X_{n+1} = s_j) = \sum_{i=1}^{k} w_i p_{ij} \quad 对 j=1,\cdots,k.$$

86

换句话说，如果在时刻 n 不同状态的概率是由概率向量 $\boldsymbol{w}$ 确定的，那么它在时刻 $n+1$ 的概率就是由概率向量 $\boldsymbol{w}P$ 确定的．结果，由此，一个有平稳转移概率链的初始概率向量是 $\boldsymbol{v}$，在时刻 $n+1$ 不同状态的概率是由概率向量 $\boldsymbol{v}P^n$ 确定的.

例 2.4.6　占用电话线数目的概率　再次考虑有五条电话线的办公室和转移矩阵 P 是由等式(2.4.2)所给出的马尔可夫链．假设在时刻 $n=1$，观察时刻的开始，一条线路都没使用的概率是 0.5，使用一条线路的概率是 0.3，使用两条线的概率是 0.2，那么初始概率向量就是 $\boldsymbol{v}=(0.5,0.3,0.2,0,0,0)$. 我们将首先决定一个时间段后，在时刻 2 恰好使用 j 条线的概率.

通过初等的计算，发现

$$\boldsymbol{v}P = (0.13,0.33,0.22,0.12,0.10,0.10).$$

因为这个概率向量的第一个分量是 0.13，在时刻 2 一条线路都没使用的概率就是 0.13；因为第二个分量是 0.33，在时刻 2 恰好使用一条线的概率是 0.33；以此类推.

下面，我们将确定在时刻 3 恰好使用 j 条线的概率.

利用等式(2.4.3)，就有

$$\boldsymbol{v}P^2 = (0.133,0.227,0.202,0.156,0.162,0.120).$$

因为这个概率向量的第一个分量是 0.133，在时刻 3 没有使用线路的概率就是 0.133；因为第二个分量是 0.227，在时刻 3 恰好使用一条线的概率是 0.227；以此类推. ◀

2.4.5 小结

马尔可夫链是一个随机过程，它给定所有过去的状态在下一时刻处于各个状态的条件概率只是通过最近的状态来依赖于过去的. 对于有有限多个状态和平稳转移概率的马尔可夫链来说，随时间而改变的转移概率可以用一个矩阵来表示，这个矩阵给出了由行标记的状态转移到由列标记的状态的概率（转移矩阵 P）. 初始概率向量 $\boldsymbol{v}$ 给出了时刻 1 该链处于各个状态的概率，转移矩阵和初始概率向量一起可以用来计算与马尔可夫链有关的所有概率．特别的，P^n 给出了经过 n 时段的转移的概率，$\boldsymbol{v}P^n$ 给出了在 $n+1$ 时刻不同状态的概率.

87

习题

1. 考虑例2.4.1中具有初始概率向量 $v=(1/2,1/2)$ 的马尔可夫链.
   a. 表示出在 $n=2$ 时刻状态概率的概率向量.
   b. 找出两步转移矩阵.
2. 假设天气只可能是晴或多云，连续的早晨天气的状况形成一个平稳转移概率的马尔可夫链，并假设这个转移矩阵如下：

	晴	多云
晴	0.7	0.3
多云	0.6	0.4

   a. 如果某一天是多云，下一天也是多云的概率是什么?
   b. 如果某一天是晴天，在以后的两天都是晴天的概率是什么?
   c. 如果某一天是多云，在以后的三天中至少有一天是晴天的概率是什么?
3. 再次考虑习题2中所描述的马尔可夫链.
   a. 如果某个星期三是晴天，在接下来的星期六是晴天的概率是什么?
   b. 如果某个星期三是多云，在接下来的星期六是晴天的概率是什么?
4. 再次考虑习题2和3的条件.
   a. 如果某个星期三是晴天，在接下来的星期六和星期日两天都是晴天的概率是什么?
   b. 如果某个星期三是多云，在接下来的星期六和星期日两天都是晴天的概率是什么?
5. 再次考虑习题2中所描述的马尔可夫链. 假设某个星期三是晴天的概率是0.2，是多云的概率是0.8.
   a. 确定下一天(星期四)将是多云的概率.
   b. 确定星期五将是多云的概率.
   c. 确定星期六将是多云的概率.
6. 假设一个学生可能准时去上课或迟到，连续几天他准时或迟到的事件形成一个具有平稳转移概率的马尔可夫链. 同样假设如果某一天他迟到，那么他在下一天准时的概率是0.8，而且，如果某一天他准时，那么他在下一天迟到的概率是0.5.
   a. 如果该学生在某一天迟到，他以后三天的每一天都准时的概率是什么?
   b. 如果该学生在某一天上课准时，他在以后三天的每一天都迟到的概率是什么?
7. 再次考虑习题6中所描述的马尔可夫链.
   a. 如果该学生在第一天上课迟到，他在第四天上课准时的概率是什么?.
   b. 如果该学生在第一天上课准时，他在第四天上课准时的概率是什么?
8. 再次考虑习题6和7的条件. 假设该学生在第一天上课迟到的概率是0.7，他准时的概率是0.3.
   a. 确定第二天上课他将准时的概率.
   b. 确定第四天上课他将准时的概率.
9. 假设一个马尔可夫链有四个状态 s_1,s_2,s_3,s_4，其平稳转移概率由下面的转移矩阵确定：

$$\begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ s_1 & 1/4 & 1/4 & 0 & 1/2 \\ s_2 & 0 & 1 & 0 & 0 \\ s_3 & 1/2 & 0 & 1/2 & 0 \\ s_4 & 1/4 & 1/4 & 1/4 & 1/4 \end{array}$$

a. 如果在给定时刻 n，链的状态是 s_3，在时刻 $n+2$ 链是 s_2 的概率什么？

88

b. 如果在给定时刻 n，链的状态是 s_1，在时刻 $n+3$ 链是 s_3 的概率什么？

10. 设 X_1 表示马尔可夫链在时刻 1 的初始状态，其转移矩阵是由习题 9 确定的，假设其初始概率如下：

$$\Pr(X_1=s_1)=1/8, \Pr(X_1=s_2)=1/4,$$
$$\Pr(X_1=s_3)=3/8, \Pr(X_1=s_4)=1/4.$$

对于以下 n 的每一个值，确定链在时刻 n 的状态 s_1，s_2，s_3 和 s_4 的概率：a. $n=2$；b. $n=3$；c. $n=4$.

11. 一位顾客购买一支牙膏，她选择品牌 A 或品牌 B，她选择与她以前所购品牌相同的概率是 1/3，选择与她以前所购品牌不同的概率是 2/3.

a. 如果她第一次购买的是品牌 A，她第五次购买品牌 B 的概率什么？

b. 如果她第一次购买的是品牌 B，她第五次购买品牌 A 的概率什么？

12. 假设有三个男孩 A、B 和 C，相互掷球，每当 A 有球时，他以 0.2 的概率扔给 B，以 0.8 的概率扔给 C，当 B 有球时，他以 0.6 的概率扔给 A，以 0.4 的概率扔给 C，当 C 有球时，他以相同的概率扔给 A 或 B.

a. 考虑这个过程是一个马尔可夫链并构造它的转移矩阵.

b. 如果在某一时刻 n，这三个男孩中的每个人有球的可能性相同，在时刻 $n+2$ 时，这个球最有可能在哪个男孩手里.

13. 假设以这样的方式反复地投掷一枚硬币，在每次投掷中出现正面和反面的可能性相同，且每次投掷都是相互独立的，以下情况除外：只要在三次连续投掷中，出现三个正面或三个反面，那么下一次投掷的结果总是相反的类型．在 $n(n\geqslant 3)$ 时刻，设这个过程的状态是由第 $n-2$，$n-1$ 和 n 次投掷的结果所确定的．证明这个过程是一个具有平稳转移概率的马尔可夫链．并构造这个转移矩阵.

14. 有两个盒子 A 和 B，每个盒子里有红球和绿球．假设盒子 A 中有一个红球和两个绿球，盒子 B 中有八个红球和两个绿球．考虑以下过程：从 A 盒中随机地选一个球，且从 B 盒中随机地选一个球．将从 A 盒中选出的球放进 B 盒，将从 B 盒中选出的球放进 A 盒，无限重复这些操作．证明：A 盒中的红球数目形成一个具有平稳转移概率的马尔可夫链，并构造马尔可夫链的转移矩阵.

15. 证明例 2.4.3 中转移矩阵中对应当前状态{AA, Aa}和{Aa, aa}的行.

16. 设在例 2.4.3 中的初始概率向量 $v=(1/16, 1/4, 1/8, 1/4, 1/4, 1/16)$，找出下一代的六个状态的概率向量.

第3章 随机变量及其分布

3.1 随机变量与离散型分布

随机变量是定义在样本空间上的实值函数，它是统计分析中对未知变量建模的主要工具. 对每个随机变量 X 和每个实数集合 A，可以计算 X 取 A 中值的概率，所有这些概率的集合就是 X 的分布. 随机变量及其分布函数主要有两类：离散型分布(本节)和连续型分布(3.2节). 离散型分布是那些对至多可数个不同值有正概率的分布. 离散型分布可以用给定每一点概率值的离散分布函数(p. f.)来表示. 称具有离散型分布函数的随机变量为离散型随机变量.

3.1.1 随机变量的定义

考虑样本空间为 S 的试验，称定义在 S 上的实值函数为随机变量，也就是说在一个特定试验中，随机变量 X 是对于样本空间 S 中的每个值 $s\in S$ 都对应一个实数值 $X(s)$ 的某个函数.

例3.1.1 抛硬币 考虑将一枚硬币抛掷10次的试验中，可以将样本空间看作有 2^{10} 种可能出现的硬币正反面的不同序列. 随机变量 X 是10次抛掷硬币中出现正面的次数. 对每种可能的包含正面和反面的序列 s，都有一个对应的 $X(s)$ 等于这一序列中正面出现的次数. 例如，如果 s 是序列：HHTTTHTTTH，那么有 $X(s)=4$. ——◀

例3.1.2 对水电的需求 考虑例1.4.2中：承包商关心一个新综合办公楼水和电的需求. 样本空间如图1-3所示. 考虑形如 (x,y) 组成的点集所表示的样本空间. 其中 x 是对水的需求，y 是对电的需求. 即对任意 $s\in S, s=(x,y)$. 对于此问题，感兴趣的一个随机变量是对水的需求. 当 $s=(x,y)$，也可以用 $X(s)=x$ 来表示. 我们感兴趣的另外一个随机变量是电的需求量 Y，当 $s=(x,y)$ 时它可以表示为 $Y(s)=y$. 我们还可以定义第三个随机变量 Z 来表征是否有至少有一种需求太高了. 令 A 和 B 是例1.4.2中的两个事件，即 A 代表用水需求至少为100的事件，B 代表用电需求至少为115的事件. 定义：

$$Z(s)=\begin{cases}1 & 若\ s\in A\cup B,\\ 0 & 若\ s\notin A\cup B.\end{cases}$$

其中，事件 $A\cup B$ 如图3-1所示. ——◀

例3.1.3 测身高 考虑以下试验：从人群中任意选择一人并以英寸为单位测其身高，那么测得的高度就是一个随机变量. ——◀

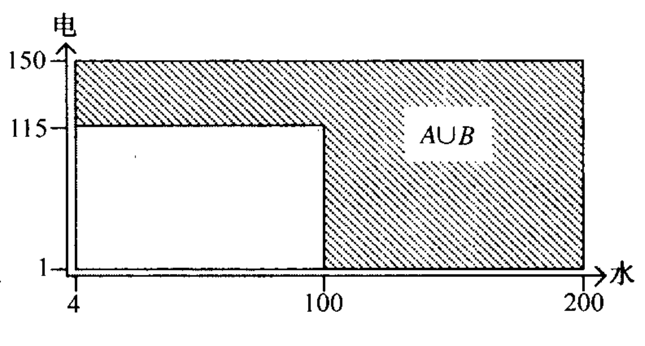


图 3-1 水电需求

3.1.2 随机变量的分布

如果对于试验的样本空间已经确定了概率分布，就可以给出所定义的随机变量 X 可能取值的概率分布. 记 A 是实轴的子集. 将 $X\in A$ 的概率记做 $\Pr(X\in A)$，则 $\Pr(X\in A)$ 等于试验的结果 s 使 $X(s)\in A$ 的概率，用符号 $\Pr(x\in A)=\Pr[s\,|\,X(s)\in A]$ 表示. 98

例 3.1.4 抛掷硬币 重新考虑掷硬币 10 次的试验，设 X 为出现正面的次数，则在这个试验中 X 可能的取值为 $0,1,2,\cdots,10$；

$$\Pr\ (X=x)\ =\binom{10}{x}\frac{1}{2^{10}},\quad x=0,1,\cdots,10.$$ ◀

例 3.1.5 水电需求 在例 1.4.2 中，我们实际上计算了例 3.1.2 中所定义的 3 个随机变量 X、Y 和 Z 的分布. 例如，设事件 A 代表用水需求至少为 100 的事件，可以表示为 $A=\{X\geqslant 100\}$ 并且 $\Pr(A)=0.5102$. 这意味着 $\Pr(X\geqslant 100)=0.5012$. 对于任意实数集 C 所对应的所有形如 $\Pr(X\in C)$ 的概率组成了 X 的分布. 可以类似于例 1.4.2 中计算 $\Pr(A)$ 那样得出 $\Pr(X\in C)$. 特别地，如果 C 是区间 $[4,200]$ 的子集，那么：

$$\Pr(X\in C)=\frac{(150-1)\times C\text{ 的长度}}{29\ 024}\tag{3.1.1}$$

例如，集合 C 是区间 $[50,175]$，其长度是 125，那么 $\Pr(X\in C)=\dfrac{149\times 125}{29\ 204}=0.6378$. 对应的样本空间的子集概率的计算如图 3-2 所示. ◀

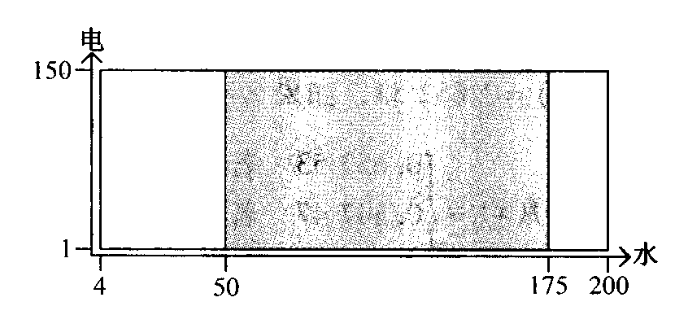


图 3-2 供水需求介于 50 与 175 之间的事件

3.1.3 离散分布

当 X 只能取有限个不同值 $x_1, x_2, \cdots, x_k$ 或者至多可列个不同值 $x_1, x_2, \cdots$，我们称随机变量 X 服从一个离散型分布或 X 是离散型随机变量. 若随机变量可以取某个区间中的任一点，称该随机变量是连续的. 关于连续型随机变量的情况将在 3.2 节中详细讨论. 如果随机变量 X 服从离散型分布，定义 X 的概率函数(简记为 p. f.)为 f：对于任意实数 x, $f(x) = \Pr(X=x)$.

显然如果随机变量 X 不能取到值 x，那么 $f(x)=0$. 并且，如果数列 $x_1, x_2, \cdots$ 包含了 X 所有可能取到的值，那么 $\sum_{i=1}^{\infty} f(x_i) = 1$. 图 3-3 给出了一个典型的概率函数图，图中的每条竖线段代表了 x 处 f 的取值 $f(x)$，这些竖线段长度之和是 1.

99

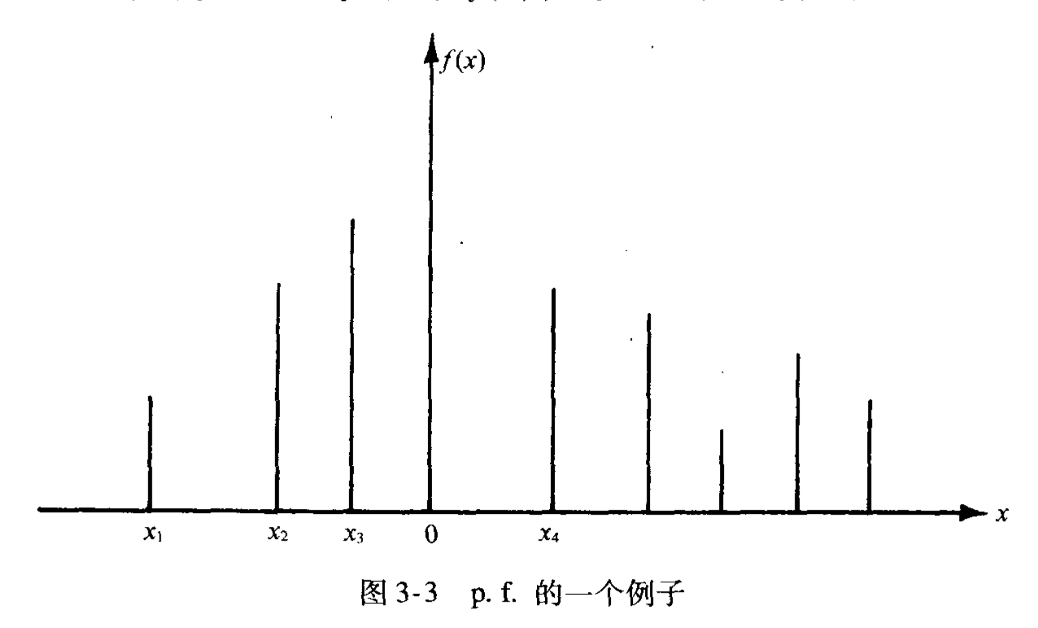


图 3-3　p. f. 的一个例子

当 X 服从离散型分布时，X 取实轴的子集 A 中值的概率定义为：

$$\Pr(x \in A) = \sum_{x_i \in A} f(x_i).$$

例 3.1.6　水电需求　考虑例 3.1.2 中的随机变量 Z，当至少有一种需求超标时，令 $Z=1$；当两种需求都低于标准值时，令 $Z=0$. 由于 Z 只能取两个不同的值，故 Z 服从离散型分布. 注意到 $\{s \mid Z(s)=1\} = A \cup B$，其中 A 和 B 如例 1.4.2 中定义，在例 1.4.2 中计算得 $\Pr(A \cup B) = 0.652\ 53$，如果 Z 有概率分布函数 f，则

$$f(x) = \begin{cases} 0.652\ 53 & \text{若 } z=1, \\ 0.347\ 47 & \text{若 } z=0, \\ 0 & \text{其他}. \end{cases}$$ ◀

一些十分常用的随机变量有特定的名称. 比如，只取两个值 0 或 1 的随机变量 Z 且 $\Pr(Z=1)=p$(如例 3.1.6 中定义的 Z,其中 $p=0.652\ 52$)称它为有参数为 p 的伯努

利分布. 下面我们给出另外两个常用的分布来结束此节.

3.1.4 整数上的均匀分布

设随机变量 X 等可能地取 $1,2,\cdots,k$ 的任一值，那么 X 的概率函数如下：

$$f(x)=\begin{cases}\dfrac{1}{k} & x=1,2,\cdots,k\\ 0 & x\neq 1,2,\cdots,k\end{cases}$$

100

称这个离散分布为整数 $1,2,\cdots,k$ 上的均匀分布. 我们常用的从 1 到 k 中随机地取一值的试验结果就服从该分布. 其中所说的“随机”意味着取到 $1,2,\cdots,k$ 中每个整数值的概率是相等的，并且这些概率之和是 1. 换言之，离散的均匀分布可以用在对象取值空间是任意有限集合的情况，但是不能被应用到取值对象为无限集合的情况，特别地，容易将均匀分布推广到任意有限个连续的整数列.

例 3.1.7 日常数 在一种公共的彩票博弈游戏中，要求参与者选择一个三位数(允许该三位数的首位是0)，于是从一只充分搅拌均匀的放满标有数字的球的大碗中随机选择三只球. 在该试验中，样本空间由所有三元数组(i_1,i_2,i_3)组成，其中 $i_j\in\{0,1,\cdots,9\}$，$j=1,2,3$. 若 $s=(i_1,i_2,i_3)$，定义 $X(s)=100i_1+10i_2+i_3$. 如 $X(0,1,5)=15$. 显然，$\Pr(X=x)=0.001$，任何 $x\in\{0,1,\cdots,999\}$，X 的分布是整数集 $0,1,2,\cdots,999$ 上的均匀分布. ◀

注： 不同随机变量可以有相同的概率分布函数. 考虑由例 3.1.7 中叙述方法得到的两列日常数，样本空间由六元数组$(i_1,\cdots,i_6)$组成，其中前三位元素是第一天得到的三个数，后三位元素是第二天得到的三个数，这六个数按取得的顺序排列. 若 $s=(i_1,\cdots,i_6)$，定义 $X_1(s)=100i_1+10i_2+i_3$，$X_2(s)=100i_4+10i_5+i_6$. 易知，X_1 和 X_2 是 s 的两个不同的函数，故不是同一个随机变量，它们可能取相等的值的概率很小. 但由于 X_1 和 X_2 取各个值的概率相等，故它们的概率分布是相同的. 又若一名商人有 1000 名顾客，他为顾客编号为 $0,1,\cdots,999$，他从这些顾客中随机地挑选一名并将其编号记为 Y. 同样地，Y 和前面所说的 X_1 和 X_2 有相同的分布但是显然 Y 和 X_1、X_2 是不同的随机变量.

3.1.5 二项分布

假设某种机器以概率 p 生产出一件次品$(0<p<1)$，以概率 $q=1-p$ 生产出一件正品. 考虑检查该机器生产出的 n 件独立的产品，并记 X 为这 n 件产品中次品的个数. 于是随机变量 X 服从离散分布，它的可能的取值为 $0,1,\cdots,n$.

这个例子与例 2.2.4 类似，对于 $x=0,1,\cdots,n$，得到含有 x 件次品 $n-x$ 件正品的一个特定排列的概率是 p^xq^{n-x}，因为共有 C_n^x 种这类可能的排列，故 $\Pr(X=x)=C_n^xp^xq^{n-x}$. X 的概率分布函数是：

101

$$f(x)=\begin{cases}C_n^x p^x q^{n-x} & \text{对 } x=0,1,\cdots,n\\ 0 & \text{其他.}\end{cases} \tag{3.1.2}$$

称由上面的分布函数表出的离散型分布为参数为 n 和 p 的二项分布. 这个分布在概率统计中十分重要，本书后面的章节将详细讨论这个分布.

本书的末尾附有二项分布取值的简表. 例如，若 X 服从参数为 $n=10$，$p=0.2$ 的二项式分布，通过查表我们可以得到 $\Pr(X=5)=0.0264, \Pr(X\geqslant5)=0.0328$.

另外再举一个二项分布的例子. 在临床实验中，假设病人从病症中痊愈的概率是 p，不能痊愈的概率是 $q=1-p$，设 Y 是 n 个互相独立的病人中痊愈的人数，则 Y 服从参数为 n 和 p 的二项分布. 事实上，由于该试验包括对 n 个相互独立的对象的观察，而对于一个病人来说只可能出现两种情况. 那么，这 n 个实验出现某种结果的概率服从参数为 n 和 p 的二项分布，其中参数 p 代表我们所要统计的结果在实验中出现的概率.

3.1.6　小结

随机变量是定义在样本空间上的实值函数. 随机变量 X 的分布是 X 落在任意实数集 A 的概率 $\Pr(X\in A)$ 的全体. 如果 X 只能取到至多可数个值，则称 X 是离散的. 这种情形下，可以用概率函数表示 X 的分布，即 $f(x)=\Pr(X=x)$ 对可能取值的集合中的 x 成立. 有些分布十分著名，因而有特别的名称. 其中之一为有限整数集上的均匀分布，更著名的一类是参数为 n 和 p 的二项式分布，其中 n 为正整数，$0<p<1$，其概率函数由(3.1.2)给出. 特别地，当 $n=1$ 时二项式分布也称为参数为 p 的伯努利分布.

习题

1. 假设随机变量 X 有整数 $10,11,\cdots,20$ 的均匀分布，试计算 X 为偶数的概率.
2. 假设随机变量 X 服从离散分布 $f(x)=\begin{cases}cx & \text{当 } x=1,\cdots,5\\ 0 & \text{其他}\end{cases}$，计算常数 c 的值.
3. 抛掷两枚均匀的骰子，并记 X 为出现两数差的绝对值. 试计算 X 的概率函数并作出其图像.
4. 将一枚均匀的硬币抛掷10次，试计算正面出现次数的分布函数.
5. 假设一个盒子中有7只红球和3只兰球，若随机且无放回地取5个球，试写出选出红球数的概率函数.　[102]
6. 设随机变量 X 服从参数为 $n=8$ 和 $p=0.5$ 的二项分布，计算 $\Pr(X<6)$.
7. 设随机变量 X 服从参数为 $n=8$ 和 $p=0.7$ 的二项分布，查书后附表计算 $\Pr(X\geqslant5)$（提示：利用 $\Pr(X\geqslant5)=\Pr(Y\leqslant3)$，其中 Y 是服从参数为 $n=8$ 和 $p=0.3$ 的二项分布）
8. 假设盒子中10%的球是红球，现从盒中随机选择20只球，每次选出后将其放回并混合均匀. 计算选出超过3只红球的概率.
9. 随机变量 X 服从离散分布：$f(x)=\begin{cases}\dfrac{c}{x^2} & x=1,2,3,\cdots\\ 0 & x\neq1,2,3,\cdots\end{cases}$，试计算常数 c 的值.

10. 某个工程师研究一个至多可以容纳7辆车的左转路口. 记随机选的一个红灯路口的停车数为X，工程师相信$X=x$的概率正比于$(x+1)(8-x)$，对$x=0,1,\cdots,7$（x的可能取值）.
    a. 写出X的概率函数.
    b. 计算$\Pr(X\geqslant 5)$

11. 证明：不存在常数c，使得函数：$f(x)=\begin{cases}\frac{c}{x} & x=1,2,3,\cdots\\ 0 & x\neq 1,2,3,\cdots\end{cases}$成为概率函数.

3.2 连续型分布

可以取到有界或者无界区间中任意值的随机变量服从连续型分布. 我们用概率密度函数(p. d. f)来刻划连续分布的随机变量，并在集合上(通常是一个区间上)对概率密度函数进行积分得到随机变量X在该集合上取值的概率.

3.2.1 概率密度函数

如果存在定义在实轴上的非负函数f，满足：对实轴的任意子集A，随机变量X落在A中的概率是f在A上的积分，那么我们说X服从连续型分布或者X是连续型随机变量. 本书中我们主要讨论A是区间的情况. 例如，对于区间$(a,b]$，

$$\Pr(a<X\leqslant b)=\int_a^b f(x)\,\mathrm{d}x. \tag{3.2.1}$$

称函数f为X的概率密度函数(probability density function)(简写为p. d. f.). 每一个概率密度函数f必须服从满足以下两个要求：

对所有x，有
$$f(x)\geqslant 0, \tag{3.2.2}$$
并且

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x=1. \tag{3.2.3}$$

103

例3.2.1 水电需求 在例3.1.5中，我们已经确定了水需求量X的分布. 从图3-2中，我们得出X所能取的最小值是4，最大值是200. 对于每一个区间$C=(c_0,c_1)\subset[4,200]$，(3.1.1)式说明：

$$\Pr(c_0<x\leqslant c_1)=\frac{149(c_1-c_0)}{29\ 204}=\frac{c_1-c_0}{196}=\int_{c_0}^{c_1}\frac{1}{196}\,\mathrm{d}x$$

于是，X的概率密度函数是：

$$f(x)=\begin{cases}\frac{1}{196} & \text{若 } 4\leqslant x\leqslant 200\\ 0 & \text{其他.}\end{cases}$$ ◀

图3-4给出了一个典型的概率密度函数. 图中，曲线下方的面积必定是1，并且

$\Pr(a < x \leqslant b)$即为图中阴影部分的面积.

注：连续分布在单个点的概率是0.（3.2.1）式的积分同样等于$\Pr(a < X < b)$，$\Pr(a \leqslant X \leqslant b)$和$\Pr(a \leqslant X < b)$. 于是，根据连续分布函数的定义可知：若$X$服从连续分布，那么对于任意的$a$，$\Pr(X=a)=0$. 正如在本书前面所说，$\Pr(x=a)=0$并不意味着$X$不可能取到$a$. 如果根据$\Pr(X=a)=0$就断言$X$不可能取到$a$的话，$X$将不能取任何值. 其实，$X$落在每个单点的概率非常小，所以我们只能在诸如区间的一个集合对其进行讨论. 这和直线在二维空间中的面积是零，但却丝毫不意味着在二维空间直线是不存在是一个道理. 图3-4中的曲线下方的两条直线面积为零，同样再一次说明了$\Pr(X=a)=\Pr(X=b)=0$.

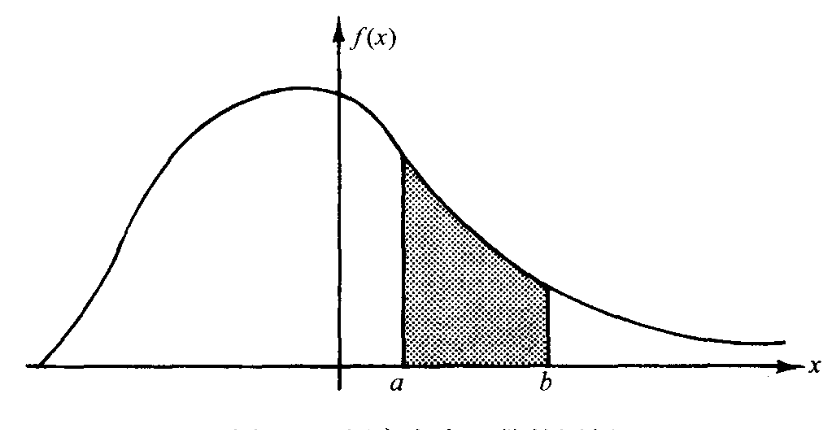


图3-4 概率密度函数的图例

3.2.2 概率密度函数的不唯一性

如果随机变量X服从连续分布，那么对于任一单点x有：$\Pr(X=x)=0$. 根据这一性质，任意一个概率密度函数的值可以在有限个甚至无限个点改变但却不影响密度函数在任意区间A上的积分. 换句话说，随机变量X的概率密度函数取值可以在任意无限点列改变但不影响涉及X的概率，即不影响X的概率分布.

104

根据前面所说的我们可以得出结论：随机变量的概率密度函数是不唯一的. 然而，在许多问题中，我们会发现这些均符合条件的概率密度函数中有一个的形式比其他概率密度函数更自然，这个函数是在整个实轴上连续的. 例如，图3-4给出了一个在整个实轴上连续的概率密度函数. 这个概率密度函数可以在一些点集上任意改变但是不影响所代表的随机变量的概率分布，但是对概率密度函数所做的该种改变必然破坏了函数的连续性，却没有对随机变量性质的研究带来任何的有利之处.

现在我们举一些例子来说明这些概念.

3.2.3 区间上的均匀分布函数

设a，b为两个给定的实数且$a < b$. 现考虑如下试验：在区间$S=$

$\{x: a\leqslant x\leqslant b\}$中取点 X，使得 X 落在 S 的任意子区间中的概率正比于该子区间的长度. 我们称随机变量 X 服从区间$[a,b]$上的均匀分布. 这里 X 表示试验的结果，该试验也常常被描述为从区间$[a,b]$中随机选取一个实数. 这里所说的"随机选取"是指点 X 取自区间中任一部分的可能性和取自该区间其他部分的可能性是相同的.

例 3.2.2 气温预报 电视天气预报报道的最高和最低温度是整数，然而这些预报是非常复杂的气象模型的结果，这些模型提供了更精确的预报，而电视预报为简单起见只把它精确到最接近的整数位. 如果要知道气象模型实际产生的温度 X，我们可以用区间$[y-1/2,y+1/2]$上的均匀分布表示温度的分布，其中 y 是电视天气预报所报道的温度. ◀

如果 X 服从区间 S 上的均匀分布，X 必须在 S 中取值，X 的概率密度函数在 S 的外部为零. 另外，由于 X 落在 S 中等长的区间的概率是相等的，而无论这些区间在什么位置，$f(x)$在 S 上必须是常值函数. 又根据

105

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x=\int_a^b f(x)\,\mathrm{d}x=1. \tag{3.2.4}$$

得出$f(x)$在 S 上的取的常值必恒为 $1/(b-a)$. X 的概率密度函数必是:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{对 } a\leqslant x\leqslant b\\ 0 & \text{其他.}\end{cases} \tag{3.2.5}$$

图 3-5 给出了概率密度函数. 例如，例 3.2.1 中的随机变量 X（水的需求量）服从$[4,200]$上的均匀分布.

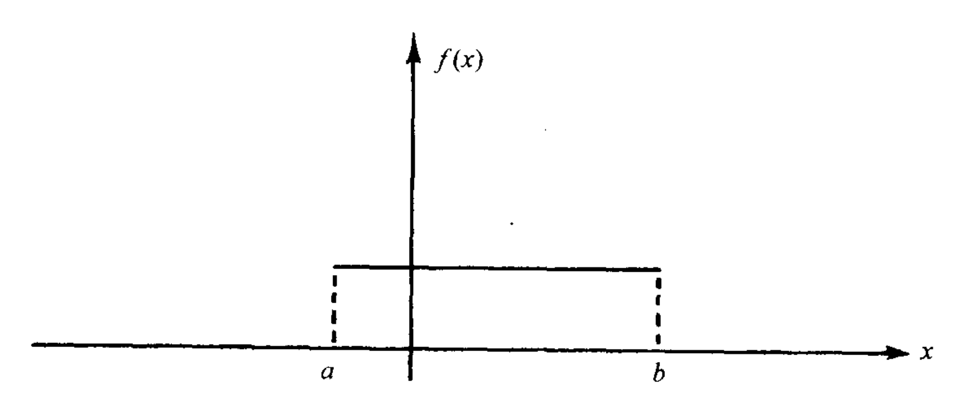


图 3-5 均匀分布的概率密度函数

(3.2.5)式说明表示在给定区间上的均匀分布的概率密度函数在该区间上取常数值，并且这一常数值就是区间长度的倒数. 由于无界区间的长度是无穷，我们不能在无界区间 $x\geqslant a$ 上定义均匀分布.

再次考虑$[a,b]$上的均匀分布. 由于在端点 a 或者 b 的概率是 0，均匀分布的区间的开闭性，即均匀分布定义在(a,b)或$(a,b]$或$[a,b]$上对概率密度函数不产生影响.

例如：随机变量 X 服从$[-1, 4]$上的均匀分布. X 的概率密度函数是:

$$f(x)=\begin{cases}\frac{1}{5} & 对于 -1\leqslant x\leqslant 4,\\ 0 & 其他.\end{cases}$$

从而

$$\Pr(0\leqslant X<2)=\int_0^2 f(x)\,dx=\frac{2}{5}.$$

3.2.4 其他连续分布

例 3.2.3 从概率密度函数计算概率 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}cx & 对 0\leqslant x\leqslant 4,\\ 0 & 其他.\end{cases}$$

106

其中 c 是给定常数. 现在我们来确定 c 的值，并由此计算 $\Pr(1\leqslant X\leqslant 2)$ 和 $\Pr(X>2)$ 的值.

对于每个概率密度函数必有 $\int_{-\infty}^{\infty} f(x)\,dx=1$. 于是，对本例有：

$$\int_0^4 cx\,dx=8c=1.$$

于是，得到 $c=1/8$. 从而，

$$\Pr(1\leqslant X\leqslant 2)=\int_1^2 \frac{1}{8}x\,dx=\frac{3}{16}$$

以及

$$\Pr(X>2)=\int_2^4 \frac{1}{8}x\,dx=\frac{3}{4}.$$ ◀

注： 计算规范化常系数. 在例 3.2.3 中第一个计算表明了简化许多统计结果的重要观点. 可以不严格地给出常数 c 的值而确定随机变量 X 的概率密度函数. 然而，我们可以通过概率密度函数的积分必为 1 这一性质计算得到 c. 这一手法十分常见，尤其是在第 7 章中我们可以找到对观测数据综述的样本分布，即除了一个常数因子外可以确定随机变量的概率密度函数. 由于概率密度函数在全空间的积分为 1，所以即便在有些情况下直接计算它的待确定常系数比较困难，我们也可以肯定的说这个常系数是唯一的.

例 3.2.4 无界随机变量 用实轴上无界区间的符合概率密度函数性质的连续型函数表示连续型随机变量是十分方便和有用的. 例如，在现实问题中某个电力系统的电压是连续分布的随机变量，它可以近似地用下面的概率密度函数表示：

$$f(x)=\begin{cases}0 & 对 x\leqslant 0,\\ \frac{1}{(1+x)^2} & 对 x>0.\end{cases}\tag{3.2.6}$$

可以证明$f(x)$满足(3.2.2)式和(3.2.3)式所要求的概率密度函数必须满足的性质.

即使电压X可能只分布在实轴上的有限区间，(3.2.6)式给出了随机变量X在它的全部值域上分布的一个很好的近似. 假设我们已知X的最大值是1000，即$\Pr(x>1000)=0$，但是用(3.2.6)假设的概率密度函数$f(x)$得到$\Pr(x>1000)=0.001$. 如果(3.2.6)在区间(0,1000)上能和随机变量X的分布充分拟合，我们用(3.2.6)作为概率密度函数比用其他符合在$x\leqslant1000$上与其只差一个常系数、在$x>1000$上取值为零的概率密度函数方便. 尤其在我们还不确定电压的最大值是1000时更应该选择(3.2.6)作为近似的概率密度函数. ◀

107

例3.2.5 无界概率密度函数 因为概率密度函数的取值是一个概率密度而非概率，这个值可以取大于1的值，甚至，我们还可以取到在$x=0$的邻域趋于无穷的概率密度函数：

$$f(x)=\begin{cases}\dfrac{2}{3}x^{-1/3} & \text{当 } 0<x<1\\ 0 & \text{其他.}\end{cases}\qquad(3.2.7)$$

即使上述函数$f(x)$无界，我们也可以证明它满足(3.2.2)和(3.2.3)所述的概率密度函数的性质. ◀

3.2.5 扩展研究：混合型分布

现实问题中所遇到的大部分分布既非离散型分布又不是连续型分布，然而我们将证明有时候很有必要考虑离散型分布和连续型分布的混合型分布.

假设例3.2.4中讨论的电力系统中用电压表来度量电压X，对电压$X<3$时可以精确地记录下来，而当$X\geqslant3$时只能记录下3. 令Y是电压表所记录的电压，那么Y的分布如下：

首先，$\Pr(Y=3)=\Pr(X\geqslant3)=1/4$. 由于$Y$取单点值3的概率是1/4，可知$\Pr(0<Y<3)=3/4$. 另外，当$0<X<3$时有$Y=X$，故可根据$X$在同一区间上的概率密度函数(3.2.6)得到(0,3)上Y的分布. 所以Y的分布由一个在区间(0,3)上的概率密度函数和一个取在单点$Y=3$的概率组成.

3.2.6 小结

一个连续分布由其概率密度函数刻划. 当对于实轴上任意一个区间(a,b)，正值函数f都满足$\Pr(a<X<b)=\int_a^b f(x)\,\mathrm{d}x$，$f$可作为$X$的概率密度函数. 连续随机变量$X$具有性质：对任意的$X=x$，$\Pr(X=x)=0$. 如果概率密度函数在区间$[a,b]$上取常数值，而在$[a,b]$外取0，我们称该分布为$[a,b]$上的均匀分布.

108

习题

1. 设随机变量 X 的概率密度函数如例(3.2.5)，试计算 $\Pr(X\leqslant 8/27)$.
2. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}\frac{4}{3}(1-x)^3 & \text{当 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

   试作出概率密度函数的图像，并计算以下概率：a. $\Pr(X<\frac{1}{2})$；b. $\Pr(\frac{1}{4}<X<\frac{3}{4})$；c. $\Pr(X>\frac{1}{3})$.
3. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}\frac{1}{36}(9-x^2) & \text{当 } -3\leqslant x\leqslant 3,\\ 0 & \text{其他.}\end{cases}$$

   试作出概率密度函数的图像，并计算以下概率(a) $\Pr(X<0)$；(b) $\Pr(-1<X<1)$；(c) $\Pr(X>2)$.
4. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}cx^2 & \text{当 } 1\leqslant x\leqslant 2,\\ 0 & \text{其他.}\end{cases}$$

   a. 计算常数 c 并作出概率密度函数的图像.

   b. 计算 $\Pr(X>\frac{2}{3})$.
5. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}\frac{1}{8}x & \text{当 } 0\leqslant x\leqslant 4,\\ 0 & \text{其他.}\end{cases}$$

   a. 求 t 值，使其满足 $\Pr(X\leqslant t)=1/4$.

   b. 求 t 值，使其满足 $\Pr(X\geqslant t)=1/2$.
6. 设随机变量 X 的概率密度函数与习题五给出的相同，在观察到随机变量 X 的值后令 Y 是最接近于 X 的整数. 计算随机变量 Y 的概率函数.
7. 设随机变量 X 服从区间$[-2,8]$上的均匀分布. 写出 X 的概率密度函数，并计算 $\Pr(0<X<7)$.
8. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}ce^{-2x} & x>0,\\ 0 & \text{其他.}\end{cases}$$

   a. 计算常数 c 并作出概率密度函数的图像.

   b. 计算 $\Pr(1<X<2)$.

9. 试证明：不存在常数 c 使下列函数成为概率密度函数：

$$f(x)=\begin{cases}\dfrac{c}{1+x} & x>0,\\ 0 & \text{其他.}\end{cases}$$

10. 随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}\dfrac{c}{(1-x)^{1/2}} & \text{当 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

a. 计算常数 c 并作出概率密度函数的图像.

b. 计算 $\Pr(X\leqslant 1/2)$.

11. 试证明：不存在常数 c 使下列函数成为概率密度函数：

$$f(x)=\begin{cases}\dfrac{c}{x} & 0<x<1,\\ 0 & x\leqslant 0, x\geqslant 1.\end{cases}$$

12. 试确定例 3.1.2 中的随机变量 Y（用电需求）的分布. 并计算 $\Pr(Y\leqslant 50)$.

3.3 分布函数

尽管离散型分布和连续型分布分别由它们的概率函数和概率密度函数所刻划，但所有的分布都可以通过分布函数来描述. 称分布函数的逆函数为分位数函数，它在表示一个分布中的概率的位置时是很有用的.

109

3.3.1 定义和基本性质

定义随机变量 X 的分布函数(distribution function) F 为如下的关于每个实数 x 的函数：

$$F(x)=\Pr(X\leqslant x)\qquad \text{当 } -\infty<x<+\infty. \tag{3.3.1}$$

有必要着重指出的是，无论该随机变量 X 的分布是离散型、连续型或是混合型的，它的分布函数都是通过这种方式定义的. 分布函数缩写为 d. f. ，有一些学者用“累积分布函数”来替代“分布函数”，并缩写为 c. d. f.

从式(3.3.1)得出，每一个随机变量 X 的分布函数是定义在实轴上的函数. 因为 $F(x)$ 表示事件 $\{X\leqslant x\}$ 的概率，对每个点 x，必存在区间 $[0,1]$ 中的一个数 $F(x)$.

进一步由式(3.3.1)，每个随机变量的分布函数必满足以下三个性质：

性质 3.3.1 函数 $F(x)$ 是 x 的单调非减函数，即，如果 $x_1<x_2$，那么 $F(x_1)\leqslant F(x_2)$.

证明： 如果 $x_1<x_2$，那么事件 $\{X\leqslant x_1\}$ 的发生也同样意味着事件 $\{X\leqslant x_2\}$ 已经发生.

因此 $\Pr\{X\leqslant x_1\}\leqslant \Pr\{X\leqslant x_2\}$ ■

图3-6就是一个分布函数的例子. 它表明在全实轴上$0 \leqslant F(x) \leqslant 1$，而且尽管$F(x)$在$x_1 \leqslant x \leqslant x_2$及$x \geqslant x_4$上为常数，当$x$增加时$F(x)$总是单调非减的.

性质3.3.2 $\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to +\infty} F(x) = 1.$

证明：事实上，当$x \to -\infty$时，$F(x) = \Pr(X \leqslant x)$必趋向于0，当$x \to +\infty$时，$\Pr(X \leqslant x)$必趋向1. 利用1.7节的习题12和13可以严格地建立这些关系. ■

图3-6给出了性质3.3.2中所指的极限值. 该图中，函数$F(x)$的取值实际上在$x = x_4$处变成1，而当$x > x_4$时保持为1，因此可以得出结论$\Pr(X \leqslant x_4) = 1$，并且$\Pr(X > x_4) = 0$. 另一方面，从图3-6的描述来看，当$x \to -\infty$，$F(x)$趋于0，但是在任意有限的x点处都达不到0. 因此，对任意有限的x，无论多么小，$\Pr(X \leqslant x) > 0$

一个分布函数并不一定是连续的. 事实上，$F(x)$可以在任意有限或至多可数的点处发生跳跃. 例如，图3-6中不连续的跳跃点发生于$x = x_1$及$x = x_3$处. 对每一个给定的x，我们令$F(x^-)$表示分布函数$F(y)$从x的左侧趋近于x的极限，即y从比x小的值趋近于x. 表示为：

110

$$F(x^-) = \lim_{\substack{y \to x \\ y < x}} F(y).$$

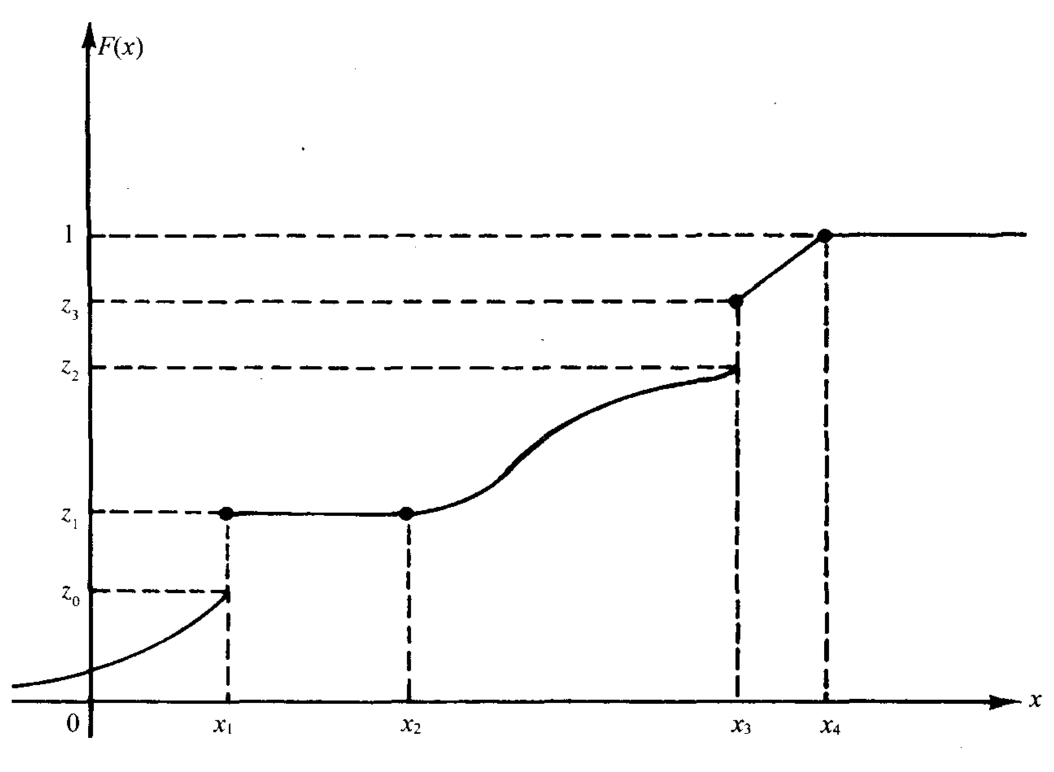


图3-6 一个分布函数的例子

类似地，我们定义$F(x^+)$为分布函数$F(y)$从x的右侧趋近于x的极限. 因此，

$$F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y).$$

如果分布函数在给定的x处连续，那么在x点处，$F(x^-) = F(x^+) = F(x)$.

性质3.3.3 分布函数总是右连续的，即在任何x点，有$F(x) = F(x^+)$.

证明：令 $y_1 > y_2 > \cdots$ 是一列递减的数，且 $\lim\limits_{n\to\infty} y_n = x$. 那么事件 $\{X \leqslant x\}$ 是所有事件 $\{X \leqslant y_n\}, n = 1,2,\cdots$ 的交. 于是，利用 1.7 节的习题 13 可以得到，

$$F(x) = \Pr(X \leqslant x) = \lim_{n\to\infty} \Pr(X \leqslant y_n) = F(x^+).$$ ■

从性质 3.3.3 可以得到，在任意发生跳跃的点 x 处，有

$$F(x^+) = F(x), F(x^-) < F(x).$$

事实上从图 3-6 中可由以下事实看到这个性质，在不连续点 $x = x_1$，$x = x_3$，$F(x_1)$，$F(x_3)$ 分别取值为 z_1，z_3.

111

3.3.2 利用分布函数确定概率

如果随机变量 X 的分布是已知的，那么 X 落在实数轴上任意一个特定区间的概率都能够由分布函数来确定. 我们将对四种不同类型的区间来进行概率的计算.

定理 3.3.1 对每个 x 值，

$$\Pr(X > x) = 1 - F(x). \tag{3.3.2}$$

证明：因为 $\Pr(X > x) = 1 - \Pr(X \leqslant x)$. 因此，利用(3.3.1)可得到(3.3.2). ■

定理 3.3.2 对所有满足 $x_1 < x_2$ 的 x_1 和 x_2，有

$$\Pr(x_1 < X \leqslant x_2) = F(x_2) - F(x_1). \tag{3.3.3}$$

证明：$\Pr(x_1 < X \leqslant x_2) = \Pr(X \leqslant x_2) - \Pr(X \leqslant x_1)$. 因此，利用(3.3.1)得到(3.3.3). ■

例如，如果 X 的分布函数在图 3-6 中所示，利用定理 3.3.1 和 3.3.2 可以得到，

$\Pr(X > x_2) = 1 - z_1$ 以及 $\Pr(x_2 < X \leqslant x_3) = z_3 - z_1$. 而且由于在区间 $[x_1, x_2]$ 上，$F(x)$ 是常值，所以 $\Pr(x_1 < X \leqslant x_2) = 0$.

在前面所有讨论中出现的关系和接下来的定理中，区分严格不等式和弱不等式是非常重要的. 如果 $F(x)$ 在给定的 x 处有一个跳跃，那么 $\Pr(X \leqslant x)$ 和 $\Pr(X < x)$ 将会不同.

定理 3.3.3 对每一个 x，有

$$\Pr(X < x) = F(x^-). \tag{3.3.4}$$

证明：令 $y_1 < y_2 < \cdots$ 是一列递增的数，且 $\lim\limits_{n\to\infty} y_n = x$. 那么，可以证明

$$\{X < x\} = \bigcup_{n=1}^{\infty} \{X \leqslant y_n\}.$$

因此，利用 1.7 节的习题 12 可以得到，

$$\begin{aligned}\Pr(X < x) &= \lim_{n\to\infty} \Pr(X \leqslant y_n) \\ &= \lim_{n\to\infty} F(y_n) = F(x^-).\end{aligned}$$ ■

例如，图 3-6 中所示的分布函数，$\Pr(X < x_3) = z_2$ 以及 $\Pr(X < x_4) = 1$.

最后，我们将证明对于每个 x，$\Pr(X = x)$ 等于 F 在点 x 的跳跃度. 如果 F 在 x 处无跳跃，则 F 连续，$\Pr(X = x) = 0$.

定理3.3.4 对于每一个x，有

[112]

$$\Pr(X=x)=F(x)-F(x^-). \tag{3.3.5}$$

证明：$\Pr(X=x)=\Pr(X\leqslant x)-\Pr(X<x)$总是成立的. 由对于每一$x$有$\Pr(X\leqslant x)=F(x)$的事实和定理3.3.3就可以得到关系式(3.3.5). ■

例如在图3-6中，$\Pr(X=x_1)=z_1-z_0$，$\Pr(X=x_3)=z_3-z_2$，而对于X的其他取值其概率值都为0.

3.3.3 离散分布的分布函数

从分布函数的定义及其性质看到，如果$a<b$，且$\Pr(a<X<b)=0$，那么$F(x)$在区间$a<x<b$上是水平的，即保持常值. 进一步，在任何使得$\Pr(X=x)>0$的点x处，分布函数将产生一个跃度为$\Pr(X=x)$的跳跃.

假设X是一个离散型分布，其概率函数为$f(x)$，相应地，它的分布函数的性质暗含了$F(x)$必有以下形式：在X可能取到的每一个x_i都有一个跃度为$f(x_i)$的跳跃；并且在连续的两个跳跃之间$F(x)$是常值. 可见，离散型随机变量的分布可以等价地由X的概率函数或者X的分布函数来表示.

3.3.4 连续分布的分布函数

现在考虑一个具有连续分布的随机变量X，令$f(x)$和$F(x)$分别表示X的概率密度函数和分布函数. 因为任意单点的概率值为0，故$F(x)$不会有跳跃. 所以$F(x)$是定义在整个实数轴上的连续函数. 进一步，由于

$$F(x)=\Pr(X\leqslant x)=\int_{-\infty}^{x}f(t)\,\mathrm{d}t, \tag{3.3.6}$$

于是在$f(x)$的任意连续点处得到，

$$F'(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x}=f(x). \tag{3.3.7}$$

故连续型随机变量X的分布可以由X的密度函数或分布函数来表示.

例3.3.1 通过分布函数来计算概率密度函数 假设在某个电力系统中，电压值X是一个随机变量，它的分布函数如下：

$$F(x)=\begin{cases}0 & \text{当}\ x<0,\\ \dfrac{x}{1+x} & \text{当}\ x\geqslant 0.\end{cases}$$

该函数满足我们在先前所述的分布函数所要求的三个性质. 进而因为在整个实轴上这个分布函数是连续的，而且除了0点外处处可微，故X的分布是连续的. 于是除了0点外，可以通过(3.3.7)获得X的密度函数. 因此可以任意给定$f(x)$在单点$x=0$处的值. 例3.2.4中，计算导数$F'(x)$时，即可以由例3.2.4的式(3.3.6)

[113]

给出$f(x)$. 相反地，如果X的密度函数由等式(3.2.6)给出，那么利用等式(3.3.6)易见$F(x)$正是由本例所给出的函数. ◀

3.3.5 分位数函数

随机变量 X 的分布函数给出了对于所有实数 x，$X\leqslant x$ 的概率. 通常我们会选定某个概率值，像 1/2，我们想要知道在分布函数的什么地方可以找到这个概率. 例如，假设 X 代表明天将会下雨的雨量，而且我们想要在 X 上下一个赌注如下：如果 $X\leqslant x_0$，可以获得 1 美元，如果 $X>x_0$，将损失 1 美元. 为了使这个赌博公平起见，就需要让 $\Pr(X\leqslant x_0)=\Pr(X>x_0)=1/2$. 当然，我们可以尝试所有的实数 x，找出使得 $F(x)=1/2$ 的点，令 x_0 等于它即可. 如果 F 是一一对应的函数，那么 F 有逆函数 F^{-1}，且 $x_0=F^{-1}(1/2)$.

1. 连续分布的分位数

当随机变量 X 的分布是连续的并且在 X 的所有的可能取值处是一一对应的，我们就称函数 F^{-1}为 X 的分位数函数. 对每一个 p，$0<p<1$，称 $F^{-1}(p)$的值为 X 的 p 分位数.

有时我们也称之为 X 的 $100p$ 百分位数. 如果令 x 表示 X 的 $100p$ 百分位数，那么就说 X 有 $100p$ 百分位数的可能落在 x 或者 x 以下. 标准化的考试分数常以百分数的形式表示，如果你的分数是第 83%，则表明有 83% 的分数等于或者小于你的分数.

例 3.3.2 风险值(Value at Risk) 投资组合管理人非常关心在固定的时间内，该组合可能会损失多少. 令 X 为一个月内给定的投资组合价值的变化. 假设 X 具有如图 3-7 的概率密度函数. 投资经理计算出风险投资领域中所熟知的风险值(用 VaR 表示). 在本例中，投资经理给出了某个概率水平，比如概率 0.99 对应的所能承受的最大损失. 这个值和 X 分布的 0.01 分位数紧密相关. 对于 0.01 分位数 x_0，有 $\Pr(X\geqslant x_0)=0.99$，于是说明有 0.99 的可能投资组合价值的变化量不小于 x_0，事实上正如图 3-7 中阴影区域所示 $x_0=-4.14$，因为损失是负的变化，所以以 0.99 的概率最大损失的值是 $-x_0=4.14$，等于概率水平为 0.99 的一个月的风险值. ◀

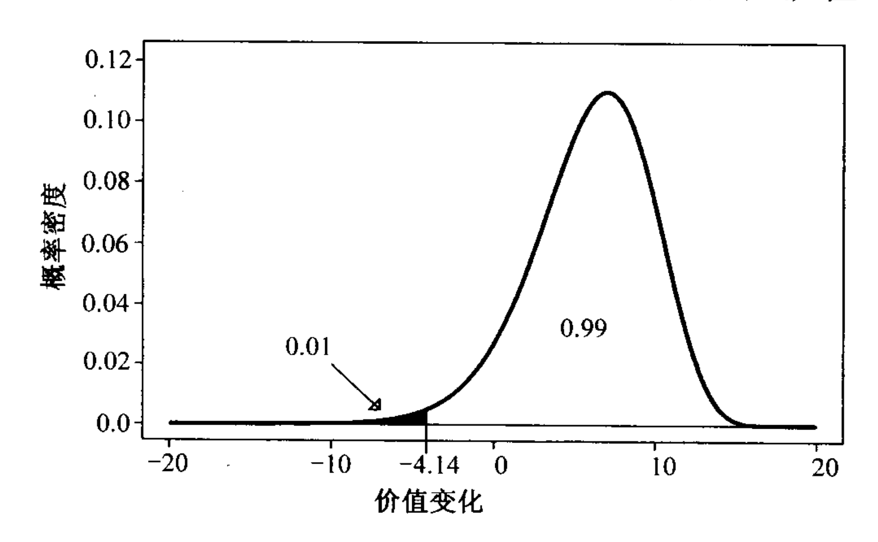


图 3-7 在低于 1% 概率投资组合价值变化量的分布密度函数

例 3.3.3 区间上的均匀分布 令 X 在区间 $[a,b]$ 上具有均匀分布. X 的分布函数为:

$$F(x)=\Pr(X\leqslant x)=\begin{cases}0 & 当\ x<a,\\ \int_a^x \frac{1}{b-a}\mathrm{d}u & 当\ a<x\leqslant b,\\ 1 & 当\ x>b.\end{cases}$$

114 上式的积分等于 $(x-a)/(b-a)$. 所以对于任意的 $x\in(a,b)$, $F(x)=(x-a)/(b-a)$, 在 X 所定义的整个区间上是严格单调增函数. 该函数的逆函数正是 X 的分位数函数, 我们使 $F(x)=p$, 并解出 x:

$$\frac{x-a}{b-a}=p,$$

$$x-a=p(b-a),$$

$$x=a+p(b-a)=pb+(1-p)a.$$

于是 X 的分位数函数是 $F^{-1}(p)=pb+(1-p)a, p\in(0,1)$. 特别地, $F^{-1}(1/2)=(b+a)/2$. ◀

某些分位数函数有特别的名称. 1/2 分位数或者第 50 的百分位数就是我们在 4.5 节中所谓的中位数. 而 1/4 或者第 25 百分位数就称为下侧四分位数. 3/4 分位数或者第 75 的百分位数就称为上侧四分位数.

注: 类似于分布函数, 分位数仅仅依赖于分布. 任意两个随机变量, 只要有相同的分布就有相同的分位数函数. 当我们涉及某个 X 的分位数之时. 指的就是 X 的分布的分位数函数.

2. 离散型分布的分位数

我们同样可以很方便地计算出离散型分布的分位数. 一般而言, 单调非减函数存在逆函数. 令 F 表示任意一个分布函数, 定义分布 F 的 p 分位数为使得满足 $F(x)\geqslant p$ 的最小的那个 x. 并记 $x=F^{-1}(p)$. 例如图 3-6, 令 $z_0\leqslant p\leqslant z_1$. 则满足 $F(x)\geqslant p$ 的最小的 x 正是 x_1. 对每一个 $x<x_1$, 我们有 $F(x)<z_0, F(x_1)=z_1$. 注意到对于任意

115 的 x_1 和 x_2 之间的所有 x, 均有 $F(x)=z_1$, 但是由于 x_1 是这些值中最小的, 所以 x_1 就是分位数. 因为分布函数是右连续的, 于是对于任意的 $0<p<1$, 满足 $F(x)\geqslant p$ 的最小的 x 是存在的. 而对于 $p=0$ 或者 $p=1$, 并不能保证这样的 x 存在. 例如图 3-6, $F(x_4)=1$, 但是却没有 x 使得 $F(x)=0$.

例 3.3.4 二项式分布的分位数 令 X 服从二项式分布, 其参数分别是 5 和 0.3. 书后面给出了 X 二项式分布的表格:

x	0	1	2	3	4	5
$f(x)$	0.1681	0.3602	0.3087	0.1323	0.0284	0.0024
$F(x)$	0.1681	0.5283	0.8370	0.9693	0.9977	1

(分布表格中有一些近似的误差)因此例如, 该分布的 0.5 分位数是 1, 1 同样也是 0.25 和 0.2 的分位数. 第 90 百分位数是 3, 3 同样也是第 95 百分位数. ◀

用分位数函数而不是分布函数来描述分布有一个好处就是，对于多元分布，比较容易通过表格的形式表示分位数函数. 这是因为无论 X 取什么数，分位数函数的定义域始终是(0,1)，分位数函数用来指出概率值的位置也是非常有用的. 例如，若想要确定分布函数的中位数的位置，即可说它落在分位数0.25和分位数0.75之间. 在7.5节中，我们将看到在考察相关数据之后如何运用分位数来给出未知量的估计.

3.3.6 小结

随机变量 X 的分布函数是 $F(x)=\Pr(X\leqslant x)$，$\forall x\in R$，它是右连续的. 如果令 $F(x^-)$ 表示 $F(y)$ 从 x 的下方逼近的极限值，那么 $F(x)-F(x^-)=\Pr(X=x)$. 连续型分布具有连续的分布函数，而且 F 在任何可微的地方，$F'(x)=f(x)$，其中 $f(x)$ 是分布密度函数. 而离散型分布所具有的分布在可能取值之间是常数，且在每个可能的 x 值处 $f(x)$ 都有跳跃. 对于任意的 $0<p<1$，分位数函数 $F^{-1}(p)$ 等于所有满足 $F(x)\geqslant p$ 的最小的那个 x.

习题

1. 假设随机变量 X 服从参数为 $p=0.7$ 的伯努利分布. 画出 X 的分布函数.
2. 假设随机变量 X 只能取值 $-2,0,1$ 和 4，相应的概率值如下：
$$\Pr(X=-2)=0.4,\Pr(X=0)=0.1,\Pr(X=1)=0.3,\Pr(X=4)=0.2.$$
画出 X 的分布函数.
3. 假设不断地投掷一枚硬币直到第一次出现头像为止，记 X 为投掷的次数，画出 X 的分布函数图像.
4. 假设随机变量 X 的分布 F 如图3-8所示，试求出下面的概率值：

[116]

a. $\Pr(X=-1)$ b. $\Pr(X<0)$ c. $\Pr(X\leqslant 0)$ d. $\Pr(X=1)$ e. $\Pr(0<X\leqslant 3)$ f. $\Pr(0<X<3)$ g. $\Pr(0\leqslant X\leqslant 3)$ h. $\Pr(1<X\leqslant 2)$ i. $\Pr(1\leqslant X\leqslant 2)$ j. $\Pr(X>5)$ k. $\Pr(X\geqslant 5)$ l. $\Pr(3\leqslant X\leqslant 4)$

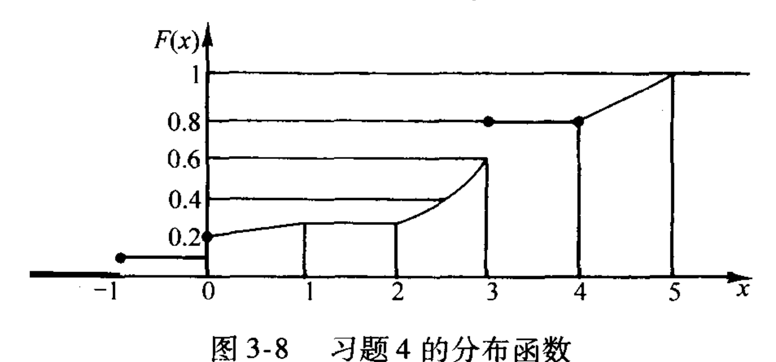


图3-8 习题4的分布函数

5. 假设随机变量 X 有如下分布函数：

$$F(X)=\begin{cases}0 & \text{当 } x\leqslant 0,\\ \frac{1}{9}x^2 & \text{当 } 0<x\leqslant 3,\\ 1 & \text{当 } x>3.\end{cases}$$

计算并画出 X 的分布函数.

6. 假设随机变量 X 有如下分布函数：

$$F(X)=\begin{cases} e^{x-3} & \text{当 } x\leqslant 3, \\ 1 & \text{当 } x>3. \end{cases}$$

计算并画出 X 的分布函数.

7. 如3.2节习题7的假设，随机变量 X 有在区间$[-2,8]$上的均匀分布. 计算并画出 X 的分布函数图像.
8. 假设在 xy 平面上，在由方程 $x^2+y^2=1$ 确定的单位圆盘内随机选一点，并且假设这个点落在该圆内某个区域的概率正比于这个区域的面积相对于单位圆面积的比例. 记随机变量 Z 为该点到圆的中心的距离. 计算并画出 Z 的分布函数.
9. 假设 X 服从$[0,5]$上的均匀分布，而定义随机变量 Y：当 $X\leqslant 1$ 时，$Y=0$，当 $X\geqslant 3$ 时，$Y=5$，其他情况下 $X=Y$. 请画出 Y 的分布函数.
10. 对例3.3.1中的分布函数，计算分位数函数.
11. 对于习题5的分布函数，确定分位数函数.
12. 对于习题6的分布函数，确定分位数函数.
13. 假设某个经纪人认为在随后的两个月内某项投资 X 的价值的变化服从区间$[-12,24]$上的均匀分布. 确定这两个月概率水平为0.95的风险值(VaR).
14. 计算确定参数为 $n=1$，$p=0.2$ 的二项式分布的分位数函数及其中位数.

117

3.4 二元分布

我们将单个随机变量的分布概念推广到两个随机变量的联合分布. 为此，我们依次引入两个离散型随机变量的联合分布函数、两个连续型随机变量的联合分布函数，以及任意两个随机变量的联合分布. 同时，也将介绍在一个离散型随机变量和一个连续型随机变量的联合混合概率函数和概率密度函数.

在例3.1.5中，我们已经找到了代表水的需求量 X 这个随机变量的分布. 但是，还有另一个感兴趣的关于电力需求的随机变量 Y. 事实上，例子中的承包商同时对这两个变量都是很关注的. 我们将两个随机变量的联合概率分布称为*二元分布*. 在本节和下两节中我们将考虑二元分布. 在3.7节中，将把这些讨论的结论推广到任意有限个随机变量的联合分布上去.

3.4.1 离散型联合分布

假设在某个试验中涉及两个随机变量 X 和 Y，每个都服从离散分布. 例如，选择某个大剧院的顾客群作为样本，令随机变量 X 表示样本中年龄大于60岁的顾客的数目，令另一个随机变量 Y 表示家住离大剧院25英里以上的顾客的数目. 如果 X 和 Y 服从取有限个数的离散型分布，那么(x,y)对也就只有有限个不同的可能取值. 另一方面，如果 X 或者 Y 可以取无限个可能值，那么(X,Y)也就有无限个可能的取值.

无论是哪一种情况，我们都说 X 和 Y 具有离散型的联合分布.

定义 X 和 Y 的联合概率函数，或联合 p. f. 为这样的函数 f，使得对于 xy 平面上的每一点 (x,y)，有

$$f(x,y)=\Pr(X=x,Y=y).$$

如果 (x,y) 不是随机变量对 (X,Y) 的取值，那么很显然 $f(x,y)=0$. 而且，由于只有至多有可列多个 (x,y) 对满足 $f(x,y)>0$，以及这些对可以反映所有的概率，故我们有：

$$\sum_{\text{所有}(x,y)} f(x,y)=1.$$

对 xy 平面上的每个子集 A，有

$$\Pr\left[(X,Y)\in A\right]=\sum_{(x,y)\in A} f(x,y).$$

例 3.4.1 二元离散分布概率表 假设随机变量 X 只能取值 1、2、3；随机变量 Y 只能取值 1、2、3 和 4. 可用下表表示 X 和 Y 的联合分布： 118

	Y			
X	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

图 3-9 中给出该概率函数. 我们将求出概率 $\Pr(X\geq 2,Y\geq 2)$ 和 $\Pr(X=1)$ 的值.

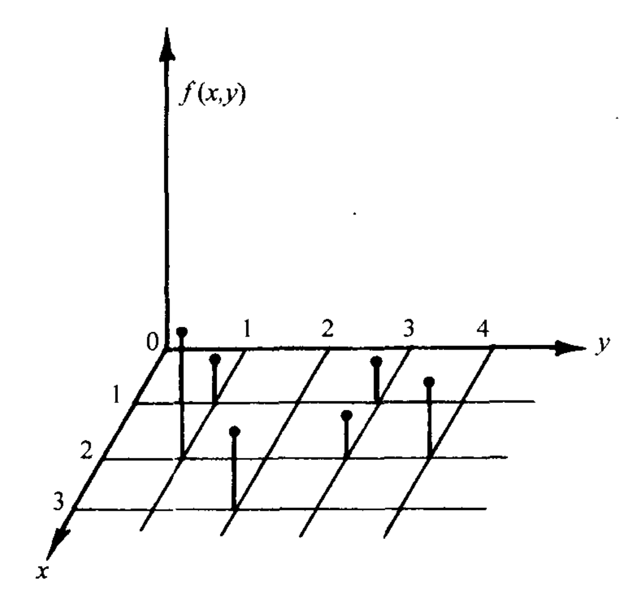


图 3-9 例 3.4.1 中 X 和 Y 的联合 p. f.

我们对于所有满足 $x\geq 2$，$y\geq 2$ 的 $f(x,y)$ 求和得到：

$$\begin{aligned}\Pr(X\geq 2,Y\geq 2)&=f(2,2)+f(2,3)+f(2,4)+f(3,2)\\&\quad+f(3,3)+f(3,4)\\&=0.5\end{aligned}$$

把表格中的第一行中的概率值求和就可以得到：

$$\Pr(X=1)=\sum_{y=1}^{4}f(1,y)=0.2.$$ ◀

注：分布中乘法法则推广的预习．在3.6节中，我们将会学习另一种非常有效的构造联合概率函数以及联合概率密度函数的方法，它是基于对条件概率的乘法法则的推广．

通过例3.4.2可以看到如何通过这种方法来实现．

119

例3.4.2　临床试验　例2.3.10中，我们在一个临床试验中观察了一系列病人．并假设了每一个病人成功的概率为 p，但是 p 未知．我们定义划分的事件为 $B_1,\cdots,B_{11}$，这里的 B_j 是事件 $p=(j-1)/10, j=1,\cdots,11$，且设 $\Pr(B_j)=1/11, \forall j$．同时也定义 E_i 表示第 i 个病人试验成功的事件，$i=1,2,\cdots$，因此对每个 i，j，有 $\Pr(E_i|B_j)=(j-1)/10$．

通过引入随机变量可以更容易理解这个试验．例如，如果第一个病人试验成功，则令 $X=1$，否则 $X=0$．即如果 E_1 发生，则 $X=1$，如果 E_1^c 发生，则 $X=0$．令 P 等于未知的参数 p 的值，也就是一个病人试验成功的概率．那么 X 和 P 都是两个离散的随机变量了．X 的可能值是0和1，P 的可能值由集合$\{0,1/10,\cdots,9/10,1\}$组成．由例2.3.4中同样理由（事实上是同样的计算方法）我们可以计算出 X 和 P 的联合概率函数．令 $F_1=E_1$，$F_0=E_1^c$．由先前定义了的$\{P=p\}=B_{10p+1}$，于是我们得到 X 和 P 的联合概率函数是：

$$\begin{aligned}f(x,p)&=\Pr(X=x,P=p)=\Pr(F_x\cap B_{10p+1})\\&=\Pr(F_x|B_{10p+1})\Pr(B_{10p+1}),\end{aligned}\tag{3.4.1}$$

对 $x=0$，1，以及 $p=0,1/10,\cdots,9/10$，1 均成立．正像例2.3.4那样，我们可以计算得：

$$\Pr(F_x|B_j)=\left(\frac{j-1}{10}\right)^x\left(1-\frac{j-1}{10}\right)^{1-x}.$$

简单地表为：

$$\Pr(F_x|B_{10p+1})=p^x(1-p)^{1-x}.\tag{3.4.2}$$

由于对每一个 p，$\Pr(B_{10p+1})=1/11)$，故结合(3.4.1)和(3.4.2)可以得到：

$$f(x,p)=\frac{1}{11}p^x(1-p)^{1-x},\tag{3.4.3}$$

对 $x=0$，1，以及 $p=0,1/10,\cdots,9/10,1$ 均成立． ◀

3.4.2　连续型联合分布

如果存在某个非负的定义在全 xy 平面上的函数 f 使得对平面的每一个子集 A 有，

$$\Pr[(X,Y)\in A]=\iint_A f(x,y)\mathrm{d}x\mathrm{d}y,$$

那么就称随机变量 X 和 Y 具有连续型联合分布. 称这个 f 为联合概率密度函数，或者简写为 X 和 Y 的联合 p. d. f.. 这样的联合 p. d. f. 必须满足如下两个条件：

$$f(x,y) \geqslant 0, \quad -\infty < x < \infty, -\infty < y < \infty$$

和

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1.$$

120

(X,Y) 对属于 xy 平面某个特定的区域的概率可以通过联合概率密度函数 $f(x,y)$ 在该区域上的积分来确定.

如果 X 和 Y 有连续型联合分布，那么下面的两个论述成立：(i) 在 xy 平面上每个单点和每个无穷点列的概率为 0. (ii) 在 xy 平面上的任意一维曲线的概率为 0. 因此 (X,Y) 落在平面上某一特定直线上的概率为 0，(X,Y) 落在某个特定圆周上的概率为 0.

图 3-10 给出了一个联合密度函数的例子. 夹在曲面 $z=f(x,y)$ 下和 xy 平面之间上的体积为 1. (X,Y) 对落在矩形 A 的概率等于图中以 A 为底的固体的体积. 而如图 3-10 所示该固体的顶正是曲面 $z=f(x,y)$ 的一部分.

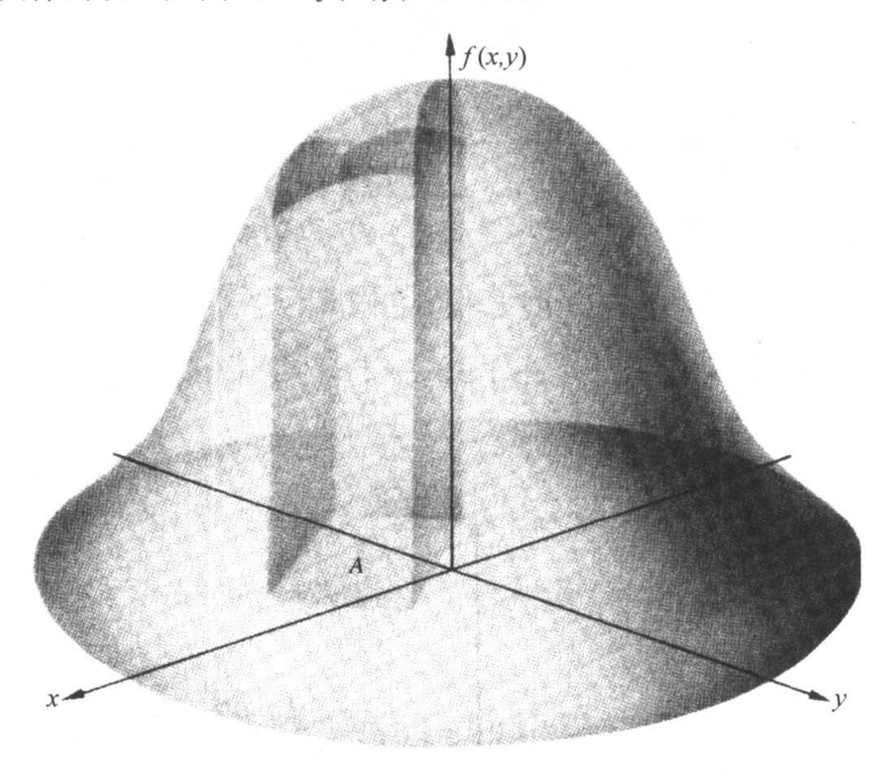


图 3-10 联合概率密度函数的例子

例 3.4.3 通过联合概率密度函数计算概率 假设 X 和 Y 的联合概率密度函数如下给出：

$$f(x,y) = \begin{cases} cx^2y & \text{当 } x^2 \leqslant y \leqslant 1, \\ 0 & \text{其他.} \end{cases}$$

121

我们首先确定常数 c，然后再求 $\Pr(X\geqslant Y)$.

图3-11描绘了满足 $f(x,y)>0$ 的点 (x,y) 的集合 S. 由于在 S 之外 $f(x,y)=0$，所以有：

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_S f(x,y)\,\mathrm{d}x\mathrm{d}y$$
$$= \int_{-1}^{1}\int_{x^2}^{1} cx^2y\ \mathrm{d}y\mathrm{d}x = \frac{4}{21}c$$

由于这个积分的值为1，故 c 值只能是21/4.

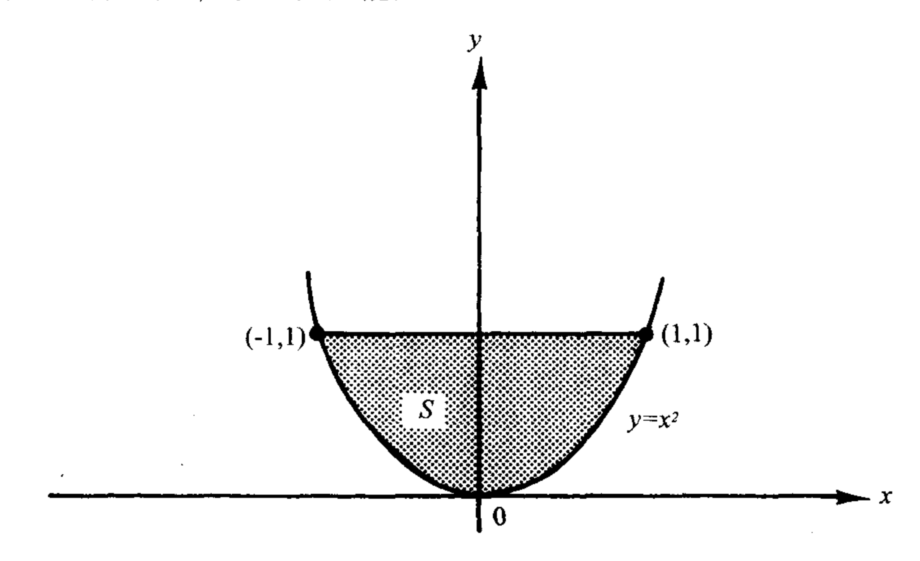
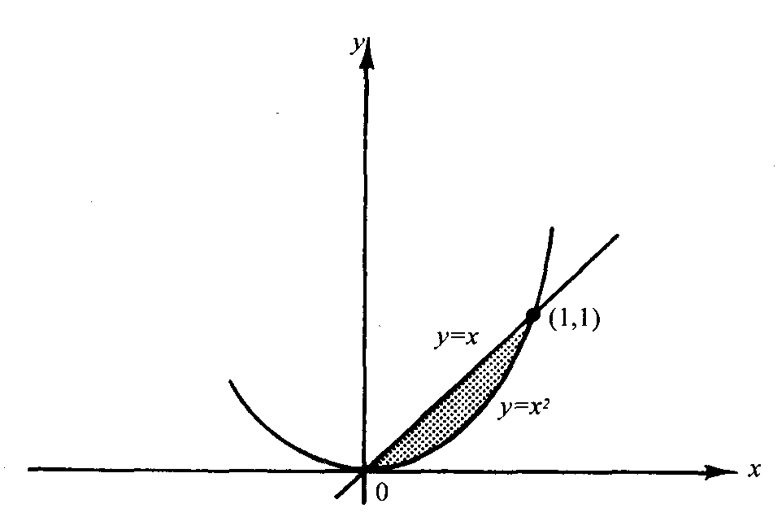


图3-11　例3.4.3中 $f(x,y)>0$ 的子集 S

图3-12表示了 S 中满足 $x\geqslant y$ 的子集 S_0. 因此，

$$\Pr(X\geqslant Y) = \iint_{S_0} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_0^1\int_{x^2}^{x} \frac{21}{4}x^2y\ \mathrm{d}y\mathrm{d}x = \frac{3}{20}.$$ ◀



122

图3-12　例3.4中满足 $x\geqslant y$，$f(x,y)>0$ 的子集 S_0

例 3.4.4 几何法确定联合概率密度函数 假设从圆盘 $x^2+y^2\leqslant 9$ 内部随机选择一点 (X,Y) 下面来确定 X 和 Y 的联合概率密度函数.

令 S 表示圆盘 $x^2+y^2\leqslant 9$ 内部的点集. 随机从该圆盘内取点等价于 X 和 Y 的联合概率密度函数在 S 上是常数，而在 S 外是 0. 因此，

$$f(x,y)=\begin{cases}c & (x,y)\in S,\\ 0 & (x,y)\notin S.\end{cases}$$

而且又必须满足

$$\iint_S f(x,y)\,\mathrm{d}x\mathrm{d}y=c\times(S\text{ 的面积})=1.$$

由于 S 的面积是 9π，故常数 c 必须是 $1/(9\pi)$. ◀

3.4.3 混合二元分布

本章中我们已经讨论了离散型和连续型的二元分布. 有时候统计学家还需要考虑混合型二元分布，随机变量中既有离散型的也有连续型的. 类似用联合概率函数来表示两个离散型随机变量的联合分布，用联合概率密度函数来表示两个连续随机变量的联合分布，那么同样也可以利用函数 $f(x,y)$ 来表示混合二元分布. (X,Y) 对落在 xy 平面的某个特定区域内的概率等于对 $f(x,y)$ 中一个变量求和并对另一个变量求积分. 我们就称这样的函数为联合概率/概率密度函数. 任意联合概率函数/概率密度函数必须满足两个条件，若 X 是离散型随机变量，可能取值 $x_1,x_2,\cdots$，Y 是连续型随机变量，那么对于任意 x，y 有 $f(x,y)\geqslant 0$ 以及

$$\int_{-\infty}^{+\infty}\sum_{i=1}^{\infty}f(x_i,y)\,\mathrm{d}y=1. \tag{3.4.4}$$

因为 f 非负，求和号和积分号可以互换，使得计算更为简便.

例 3.4.5 联合概率函数/概率密度函数 假设 X 和 Y 的联合概率/概率密度函数是：

$$f(x,y)=\frac{xy^{x-1}}{3},\quad x=1,2,3,\quad 0<y<1$$

123

我们必须验证该函数满足(3.4.4). 先对 y 积分是简便的，故计算得：

$$\sum_{x=1}^{3}\int_0^1\frac{xy^{x-1}}{3}\mathrm{d}y=\sum_{x=1}^{3}\frac{1}{3}=1.$$

假设我们需要求 $Y\geqslant 1/2$，$X\geqslant 2$ 的概率，那么就有：

$$\sum_{x=2}^{3}\int_0^1\frac{xy^{x-1}}{3}\mathrm{d}y=\sum_{x=2}^{3}\frac{1-(1/2)^x}{3}=0.5417.$$ ◀

在实际的问题中，混合分布的情况会相当复杂. 例如，X 和 Y 表示电力系统中两个特定部分出故障的时刻. 它们可能同时失效的概率为 $p,(0<p<1)$，它们在不同时刻失效的概率是 $1-p$. 进一步，若同时失效，那么它们同时失效的时刻也许就服从某个特定概率密度函数 $f(x)$；若在不同时刻失效，也许这些时刻就服从某个二元联

合概率密度函数 $g(x,y)$.

本例中 X 和 Y 的联合分布不连续：因为对任意连续分布，落在直线 $x=y$ 上的概率必须是0，而在本例中该概率是 p.

3.4.4 二元分布函数

定义两个随机变量的联合分布函数或联合 d. f. 为如下的函数 F:

$$F(x,y)=\Pr(X\leqslant x,Y\leqslant y),\quad -\infty<x<\infty,-\infty<y<\infty.$$

若任意两个随机变量 X 和 Y 的联合分布函数是 F，那么能够从 F 中得到 (X,Y) 对落在 xy 平面内特定的矩形内的概率：对于给定的 $a<b$，$c<d$,

$$\begin{aligned}&\Pr(a<X\leqslant b,c<Y\leqslant d)\\&=\Pr(a<X\leqslant b,Y\leqslant d)-\Pr(a<X\leqslant b,Y\leqslant c)\\&=[\Pr(X\leqslant b,Y\leqslant d)-\Pr(X\leqslant a,Y\leqslant d)]\\&\quad-[\Pr(X\leqslant b,Y\leqslant c)-\Pr(X\leqslant a,Y\leqslant c)]\\&=F(b,d)-F(a,d)-F(b,c)+F(a,c).\end{aligned}\tag{3.4.5}$$

因此，落入图3-13中矩形 A 中的概率就可以由上述推导出的 F 给出. 必须指出的是矩形的两条边包含在集合 A 中，而另外两条边不包含在内. 因此，若在 A 的边界上的某个点或者某条线上有正概率，那么区分公式(3.4.5)中的弱不等式和严格不等式是重要的.

124

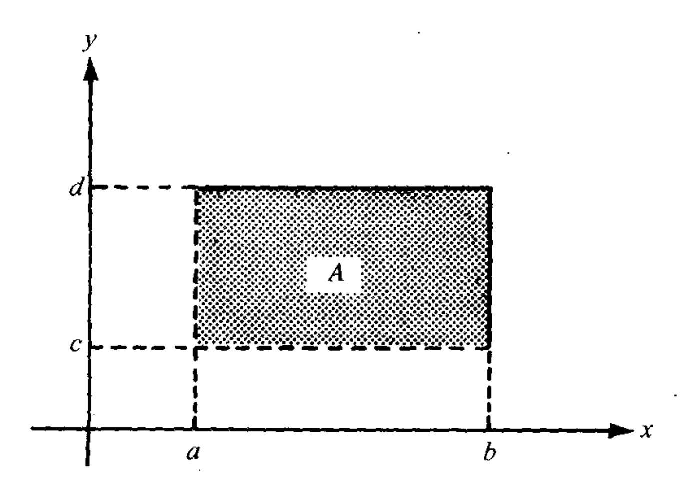


图3-13 矩形的概率

由联合分布 F 可以计算单个随机变量 X 的分布函数:

$$\begin{aligned}F_1(x)&=\Pr(X\leqslant x)=\lim_{y\to\infty}\Pr(X\leqslant x,Y\leqslant y)\\&=\lim_{y\to\infty}F(x,y)\qquad -\infty<x<\infty.\end{aligned}$$

类似地，如果记 F_2 为 Y 的分布函数，那么:

$$F_2(y)=\lim_{x\to\infty}F(x,y).$$

在下一章节将给出其他涉及 X 的单个分布,Y 的单个分布以及 X 和 Y 的联合分布之间的关系.

最后，若 X 和 Y 具有连续的联合分布，且联合概率密度函数是 f，那么在点

(x,y)的联合分布函数是：

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(r,s)\,\mathrm{d}r\mathrm{d}s.$$

这里，记号 r 和 s 是积分盲变量. 在任意二阶导数存在的点由联合分布函数求导可以得到联合概率密度函数：

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

例 3.4.6 用联合分布确定联合概率密度函数 设 X 和 Y 是随机变量，取值区间分别是$[0,2]$和$[0,2]$，且假设它们的联合分布函数如下：

$$F(x,y) = \frac{1}{16}xy(x+y) \qquad 0 \leqslant x \leqslant 2, 0 \leqslant y \leqslant 2 \tag{3.4.6}$$

125

我们首先确定随机变量 X 的分布函数 F_1，然后确定 X 和 Y 的联合概率密度函数f.

对在 xy 平面上不代表 X 和 Y 的可能取值的任意点对(x,y)，可以通过公式(3.4.6)以及式 $F(x,y)=\Pr(X \leqslant x, Y \leqslant y)$来计算 $F(x,y)$的值. 因此，若 $x<0$ 或者 $y<0$，那么 $F(x,y)=0$. 如果 $x>2$ 且 $y>2$ 那么 $F(x,y)=1$. 若 $0 \leqslant x \leqslant 2, y>2$，那么 $F(x,y)=F(x,2)$，且由(3.4.6)得：

$$F(x,y) = \frac{1}{8}x(x+2).$$

类似地，若 $0 \leqslant y \leqslant 2$，$x>2$，那么

$$F(x,y) = \frac{1}{8}y(y+2).$$

如此，在 xy 全平面上函数 $F(x,y)$都有了定义.

令 $y \to \infty$，我们便可以确定随机变量 X 的分布函数是：

$$F_1(x) = \begin{cases} 0 & x<0, \\ \frac{1}{8}x(x+2) & 0 \leqslant x \leqslant 2, \\ 1 & x>2. \end{cases}$$

进一步地，对于 $0<x<2$，$0<y<2$，

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{1}{8}(x+y).$$

同时，对 $x<0$，$y<0$，$x>2$，或者 $y>2$，

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = 0.$$

因此，X 和 Y 的联合概率密度函数是：

$$f(x,y) = \begin{cases} \frac{1}{8}(x+y) & 0<x<2, 0<y<2, \\ 0 & \text{其他}. \end{cases}$$

◀

3.4.5 小结

两个随机变量 X 和 Y 的联合分布函数是 $F(x,y)=\Pr(X\leqslant x,Y\leqslant y)$. 两个连续型随机变量的联合概率密度函数是非负函数 f 且使得 (X,Y) 对落在集合 A 中的概率等于 $f(x,y)$ 在集合 A 中的积分. 联合概率密度函数是联合分布函数的关于两个变量的二阶混合偏导数. 两个离散随机变量的联合概率函数也是非负函数 f，使得 (X,Y) 对落在集合 A 中的概率等于集合 A 中所有点上 $f(x,y)$ 的和. 联合分布函数在至多可列个 (X,Y) 对上严格大于0. 而离散型随机变量 X 和连续型随机变量 Y 的联合概率函数/概率密度函数是非负函数 f，且使得 (X,Y) 对落在集合 A 中的概率等于先对所有 $(x,y)\in A$ 的 x 求和然后对 y 求积分.

126

习题

1. 假设两个随机变量 (X,Y) 的联合概率密度函数在矩形 $0\leqslant x\leqslant 2$，$0\leqslant y\leqslant 1$ 上是常数，且在该矩形外为0.

   a. 确定在矩形上概率密度函数的常数值.

   b. 确定 $\Pr(X\geqslant Y)$.

2. 假设某个电子信号显示屏的第一排有三个灯泡且第二排有四个灯泡. 令 X 表示在 t 时刻第一排烧坏的灯泡个数，Y 表示同一时刻 t 第二排烧坏的灯泡个数. 假设 X 和 Y 的联合概率函数如下表所示：

	Y				
X	0	1	2	3	4
0	0.08	0.07	0.06	0.01	0.01
1	0.06	0.10	0.12	0.05	0.02
2	0.05	0.06	0.09	0.04	0.03
3	0.02	0.03	0.03	0.03	0.04

   确定如下的概率：

   a. $\Pr(X=2)$ b. $\Pr(Y\geqslant 2)$ c. $\Pr(X\leqslant 2,Y\leqslant 2)$ d. $\Pr(X=Y)$ e. $\Pr(X>Y)$

3. 假设 X 和 Y 有离散的联合分布，它们的联合概率函数定义如下：

$$f(x,y)=\begin{cases}c\,|x+y| & x=-2,-1,0,1,2 \text{ 且 } y=-2,-1,0,1,2,\\ 0 & \text{其他.}\end{cases}$$

   确定(a)常数 c;(b) $\Pr(X=1,Y=-2)$；(c) $\Pr(X=2)$；(d) $\Pr(|X-Y|\leqslant 1)$.

4. 假设 X 和 Y 有连续的联合分布，它们的联合概率密度函数定义如下：

$$f(x,y)=\begin{cases}cy^2 & 0\leqslant x\leqslant 2,0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

   确定(a)常数 c;(b) $\Pr(X+Y>2)$；(c) $\Pr(Y<1/2)$；(d) $\Pr(X\leqslant 1)$；(e) $\Pr(X=3Y)$.

5. 假设两个随机变量 X 和 Y 的联合概率密度函数是：

$$f(x,y)=\begin{cases}c(x^2+y) & \text{当}\,0\leqslant y\leqslant 1-x^2,\\ 0 & \text{其他}.\end{cases}$$

确定 (a) 常数 c；(b) $\Pr(0\leqslant X\leqslant 1/2)$；(c) $\Pr(Y\leqslant X+1)$；(d) $\Pr(Y=x^2)$.

6. 假设从 xy 平面内的某个区域 S 中随机选取点(X,Y)，该区域包含满足以下关系的点：$x\geqslant 0$，$y\geqslant 0$，$4y+x\leqslant 4$.

   a. 确定 X 和 Y 的联合概率密度函数.

   b. 假设 s_0 是 S 中面积为 α 的子集，确定 $\Pr((X,Y)\in S_0)$.

7. 假设从 xy 平面内的某个区域 S 中随机选取点(X,Y)，该区域中的点满足：$0\leqslant x\leqslant 1$，$0\leqslant y\leqslant 1$. 被选中位于顶点$(0,0)$，$(1,0)$，$(0,1)$，$(1,1)$的概率是分别是 0.1，0.2，0.4，0.1. 且被选中的点不是这四个顶点时为区域的内点，它们被选中的概率密度函数是一个常数.

   确定 (a) $\Pr(X\leqslant 1/4)$；(b) $\Pr(X+Y\leqslant 1)$.

8. 假设 X 和 Y 是两个随机变量，使得点(X,Y)必属于 xy 平面内由 $0\leqslant x\leqslant 3$;$0\leqslant y\leqslant 4$ 围成的矩形 S，且定义 X 和 Y 在这个矩形中的每个点(x,y)的联合分布函数如下：

$$F(x,y)=\frac{1}{156}xy(x^2+y).$$

   确定 (a) $\Pr(1\leqslant X\leqslant 2,1\leqslant Y\leqslant 2)$；(b) $\Pr(2\leqslant X\leqslant 4,2\leqslant Y\leqslant 4)$；(c) Y 的分布函数；(d) X 和Y 的联合概率密度函数；(e) $\Pr(Y\leqslant X)$.

9. 确定例 3.1.2 中需水量 X 和需电量 Y 的联合概率密度函数. 提示：参考例 1.4.2 的计算.

10. 令 Y 是到达交换台的呼叫频率（次数/小时）. 令 X 是两个小时内总的呼叫次数. 通常(X,Y)的联合概率函数/概率密度函数的一个选择是：

$$f(x,y)=\begin{cases}\dfrac{(2y)^x}{x!}e^{-3y} & y>0,\text{且}\,x=0,1,\cdots,\\ 0 & \text{其他}.\end{cases}$$

   a. 验证 f 确实是联合概率/概率密度函数. 提示：首先对利用熟知的幂级数 e^{2y} 的展开式对 x 求和.

   b. 求 $\Pr(X=0)$.

127

3.5 边际分布

在本章前面，我们介绍了随机变量的分布，在 3.4 节我们讨论了二维随机变量联合分布的一般情形. 我们经常从两个随机变量的联合分布开始，而需要他们两个中仅仅一个的分布. 称这种从联合分布计算得到的关于随机变量 X 的分布为 X 的边际分布. 每个随机变量都有边际分布函数和边际概率密度函数或概率函数，我们也引入独立随机变量的概念，它是独立事件自然的推广.

3.5.1 推导边际概率函数或边际概率密度函数

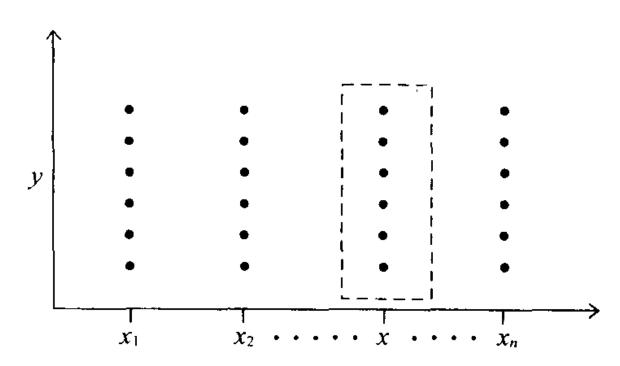
在 3.4 节可见，若已知两个随机变量 X 和 Y 的联合分布函数 F，那么能由 F 推导

出随机变量 X 的分布函数 F_1. 在这个意义下，如果 X 的分布是由 X 和 Y 的联合分布导出的，那么称 F_1 为 X 的边际分布. 类似地，如果 X 和 Y 的联合概率函数，联合概率密度函数或联合概率函数/概率密度函数 f 是已知的，那么能由 f 推导出每个随机变量的边际概率函数或边际概率密度函数.

例如，如果 X 和 Y 有离散型联合分布，其联合概率函数为 f，那么能如下得到 X 的边际概率函数 f_1：

$$f_1(x)=\Pr(X=x)=\sum_y \Pr(X=x\text{ 和 }Y=y)=\sum_y f(x,y). \tag{3.5.1}$$

换言之，对任意 x，将 $f(x,y)$ 对 Y 所有可能的取值 y 求和可以得到 $f_1(x)$. 我们通过图3-14解释其正确性. 图中，虚线框中的点集是关于第一个坐标 x 的点集对. 可以用虚线框中的事件对的并来表示事件 $\{X=x\}$，即对所有可能的 y，$A_y=\{X=x,Y=y\}$. 事件 A_y 不相交，且 $\Pr(A_y)=f(x,y)$，因为 $\Pr(X=x)=\sum_y \Pr(A_y)$ 所以等式(3.5.1)成立.



128

图3-14　通过联合概率函数计算 $f_1(x)$

类似地，能由下述关系得到 Y 的边际概率函数 f_2

$$f_2(y)=\sum_x f(x,y). \tag{3.5.2}$$

例3.5.1　由概率表推导边际概率函数　假设 X 和 Y 的联合概率函数如例3.4.1的表所示. 能通过对表中的每行求和得到 X 的边际概率函数. 通过这种方法我们可以得到 $f_1(1)=0.2$，$f_1(2)=0.6$，$f_1(3)=0.2$ 和对所有其他的 x，$f_1(x)=0$. ◀

若 X 和 Y 有连续型联合分布，联合概率密度函数为 f，那么再次由(3.5.1)式的方法确定 X 的边际概率密度函数 f_1，但现在对 Y 的所有可能值求和被对 Y 的所有可能值积分所取代. 因此，

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\quad 对\ -\infty<x<\infty.$$

类似地，由(3.5.2)式确定 Y 的边际概率密度函数 f_2，但求和再次被积分取代. 因此，

$$f_2(x)=\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\quad 对\ -\infty<y<\infty.$$

例 3.5.2 推导边际概率密度函数 假设 X 和 Y 的联合概率密度函数如例 3.4.3 中所示，我们将先确定 X 的边际概率密度函数 f_1，接着确定 Y 的边际概率密度函数 f_2.

由图 3-11 可见，X 不能取区间 $[-1,1]$ 外的任何值，所以对于 $x<-1$ 或 $x>1$ 有 $f_1(x)=0$，进一步，对 $-1\leqslant x\leqslant 1$，由图 3-11 可见 $f(x,y)\neq 0$ 当且仅当 $x^2\leqslant y\leqslant 1$. 因此，对于 $-1\leqslant x\leqslant 1$，

$$f_1(x)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}y=\int_{x^2}^{1}\left(\frac{21}{4}\right)x^2y\,\mathrm{d}y=\left(\frac{21}{8}\right)x^2(1-x^4).$$

X 的边际概率密度函数如图 3-15 所示.

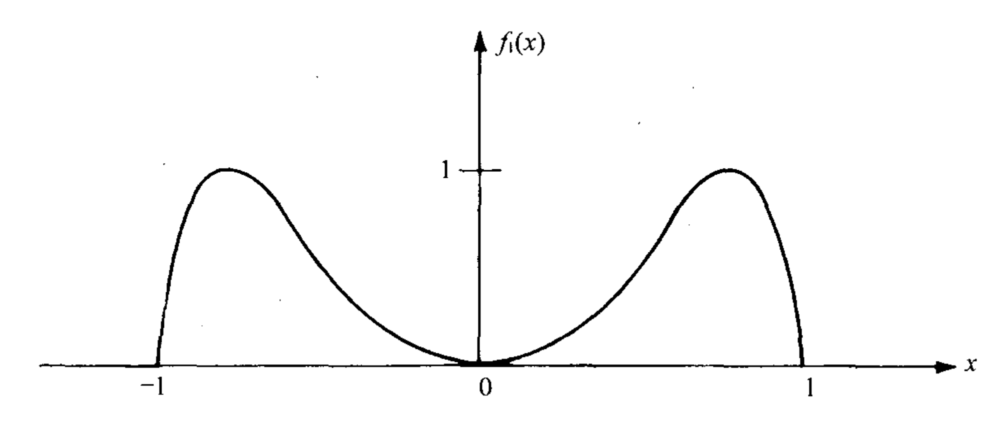


图 3-15 例 3.5.2 中 X 的边际概率密度函数

129

由图 3-11 可见，Y 不能取区间 $[0,1]$ 外的任何值. 因此，对于 $y<0$ 或 $y>1$ 有 $f_2(y)=0$. 进一步，对于 $0\leqslant y\leqslant 1$，由图 3-11 可见，$f(x,y)\neq 0$ 当且仅当 $-\sqrt{y}\leqslant x\leqslant\sqrt{y}$. 因此，对于 $0\leqslant y\leqslant 1$，

$$f_2(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x=\int_{-\sqrt{y}}^{\sqrt{y}}\left(\frac{21}{4}\right)x^2y\,\mathrm{d}x=\left(\frac{7}{2}\right)y^{5/2}.$$

Y 的边际概率密度函数如图 3-16 所示. ◀

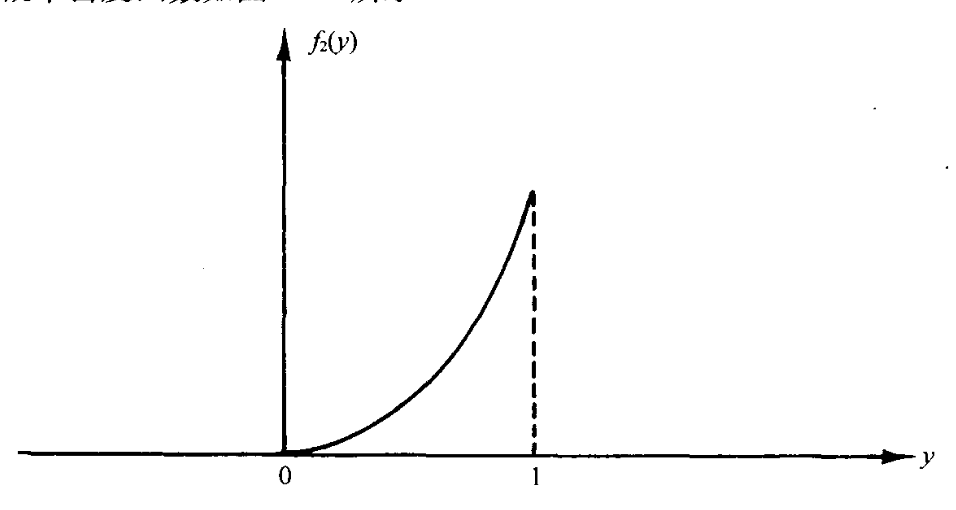


图 3-16 例 3.5.2 中 Y 的边际概率密度函数

若 X 有离散型分布，Y 有连续型分布，我们能用同样的方法由联合概率函数/概率密度函数来推导出 X 的边际概率函数和 Y 的边际概率密度函数，就如同我们由联合概率函数或联合概率密度函数推导出边际概率函数和边际概率密度函数一般. 令 f 是 X 和 Y 的联合概率函数/概率密度函数. 那么 X 的边际概率函数为

$$f_1(x) = \Pr(X = x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y, \text{对所有的 } x.$$

Y 的边际概率密度函数为

$$f_2(y) = \sum_{y} f(x,y), \text{对} -\infty < y < \infty.$$

例3.5.3　通过联合概率函数/概率密度函数确定边际概率函数和边际概率密度函数　假设 X 和 Y 联合概率函数/概率密度函数如例3.4.5中所示. 通过如下积分得到 X 的边际概率函数

$$f_1(x) = \int_0^1 \frac{xy^{x-1}}{3}\,\mathrm{d}y = \frac{1}{3},$$

对于 $x = 1,2,3$. 通过如下求和得到 Y 的边际概率密度函数

130

$$f_2(y) = \frac{1}{3} + \frac{2y}{3} + y^2, \text{对于 } 0 < y < 1.$$ ◀

尽管能通过 X 和 Y 的联合分布推导出他们的边际分布，但没有额外信息而仅通过 X 和 Y 的边际分布重构联合分布是不可能的. 例如，图3-15和图3-16所绘的边际概率密度函数没有揭示任何关于 X 和 Y 之间关系的信息. 事实上，由定义可知，X 的边际分布确定了 X 的概率而不包含任何其他随机变量的值. 能用另一个例子来进一步解释这种边际概率密度函数的性质.

假设有一枚一分硬币和一枚五分硬币，每个都掷 n 次，考虑如下的 X 和 Y 的两个定义：(i) X 是一分硬币正面朝上的次数，Y 是五分硬币正面朝上的次数. (ii) X 和 Y 都是一分硬币正面朝上的次数，即随机变量 X 和 Y 是完全相同的.

在情况(i)中，X 的边际分布和 Y 的边际分布都是同样的二项式分布. 在情形(ii)中也能得到同样的 X 和 Y 的边际分布. 然而，两个情形中 X 和 Y 的联合分布不相同. 在情形(i)中，X 和 Y 的取值完全无关，而在情形(ii)中，X 和 Y 的取值必然相同.

3.5.2　独立随机变量

称随机变量 X 和 Y 为是*独立的*，如果对于任意两个实数集 A 和 B 有，

$$\Pr(X \in A \text{ 且 } Y \in B) = \Pr(X \in A)\Pr(Y \in B). \tag{3.5.3}$$

换言之，令 A 是发生与否仅依赖于 X 的取值的任意事件，令 B 是发生与否仅依赖于 Y 的取值的任意事件. 那么 X 和 Y 是独立变量当且仅当对于所有的事件 A 和 B，A 和 B 是独立事件.

若 X 和 Y 是独立的，那么对于所有的实数 x 和 y，必然有下式成立

$$\Pr(X\leqslant x \text{ 且 } Y\leqslant y)=\Pr(X\leqslant x)\Pr(Y\leqslant y). \tag{3.5.4}$$

进一步，因为对于所有在(3.5.3)式中出现 X 和 Y 的概率类型，能由(3.5.4)式中出现的概率类型导出，从而若所有的 x 和 y 的值都满足(3.5.4)式，则 X 和 Y 必然是独立的. 这个命题的证明超出了本书的范围，所以略去. 换言之，两个随机变量 X 和 Y 是独立的当且仅当对于所有的 x 和 y 的值(3.5.4)式成立.

若记 X 和 Y 的联合分布函数为 F，X 的边际分布函数为 F_1，Y 的边际分布函数为 F_2，那么可将先前的结论重述为：**两个随机变量 X 和 Y 是独立的当且仅当对所有的实数 x 和 y 有**

$$F(x,y)=F_1(x)F_2(y).$$

假设 X 和 Y 有离散型联合分布或连续型联合分布或混合型联合分布，其联合概率函数，联合概率密度函数或联合概率函数/概率密度函数是 f. 如同以前，我们记 X 和 Y 的边际概率函数或边际分布函数为 f_1 和 f_2. 那么由上面的关系式我们得到：X 和 Y 是独立的当且仅当对于所有的实数 x 和 y，下述因式分解成立： [131]

$$f(x,y)=f_1(x)f_2(y). \tag{3.5.5}$$

如同3.2节所述，对于连续分布的密度函数可以任意改变在任意可数点集上的值. 因此，对于这样的分布而言，更准确的说法是，随即变量 X 和 Y 是独立的当且仅当能找到合适的概率密度函数 f，f_1，f_2 使得(3.5.5)式对于 $-\infty<x<\infty$ 和 $-\infty<y<\infty$ 成立.

例3.5.4 计算涉及独立随机变量的概率 假定 X 和 Y 是在特定时间内对特定地区下雨量的两个独立测量. 每个测量的概率密度函数 g 如下

$$g(x)=\begin{cases}2x & \text{对于 } 0\leqslant x\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

我们将确定 $\Pr(X+Y\leqslant 1)$ 的值.

因为 X 和 Y 是独立的，且概率密度函数均为 g，由(3.5.5)式可知，对于所有的 x 和 y，X 和 Y 的联合密度函数 $f(x,y)$ 的值能由关系 $f(x,y)=g(x)g(y)$ 确定. 因此，

$$f(x,y)=\begin{cases}4xy & \text{对于 } 0\leqslant x\leqslant 1 \text{ 且 } 0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

$x-y$ 平面的集合 $S=\{(x,y)\mid f(x,y)>0\}$ 和子集 $S_0=\{(x,y)\mid x+y\leqslant 1,(x,y)\in S\}$ 如图3-17所示. 因此

$$\Pr(X+Y\leqslant 1)=\int_{S_0}\int f(x,y)\,dxdy=\int_0^1\int_0^{1-x}4xy\,dydx=\frac{1}{6}.$$ ◀

假设 X 和 Y 有离散型分布，X 能取值 $1,2,\cdots,r$；Y 能取值 $1,2,\cdots,s$；且有

$$\Pr(X=i \text{ 且 } Y=j)=p_{ij} \quad \text{对于 } i=1,\cdots,r \text{ 和 } j=1,\cdots,s.$$

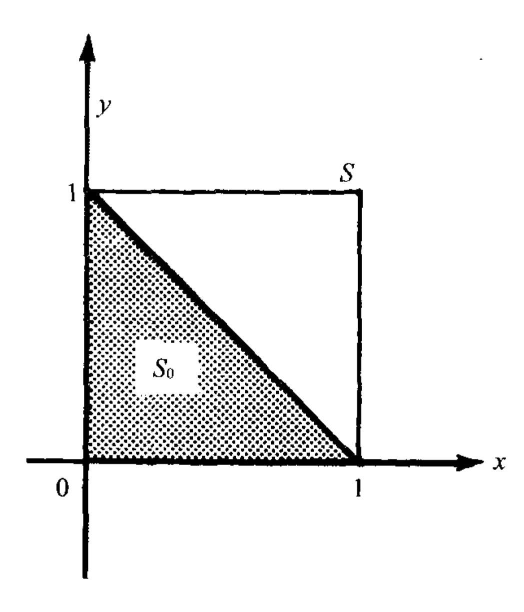


图3-17　例3.5.4中的子集 $S_0=\{(x,y)\in S\mid x+y\leq 1\}$

因此,对于 $i=1,\cdots,r$，令

$$\Pr(X=i)=\sum_{j=1}^{s}p_{ij}=p_{i+},$$

同样，对于 $j=1,\cdots,s$，令

132

$$\Pr(Y=j)=\sum_{i=1}^{r}p_{ij}=p_{+j}.$$

因此，X 和 Y 独立当且仅当下述关系对于所有的 i 和 j 成立：

$$p_{ij}=p_{i+}p_{+j}. \tag{3.5.6}$$

例3.5.5　由概率表判定随机变量是否独立　假设 X 和 Y 的联合概率函数如例3.4.1的表中所示．我们来判定 X 和 Y 是否是独立的．

在(3.5.6)式中，p_{ij}是表中第 i 行第 j 列的概率．p_{i+}是第 i 行概率的和，p_{+j}是第 j 列概率的和．由表可得 $p_{11}=0.1$，$p_{1+}=0.2$，$p_{+1}=0.4$．因此 $p_{11}\neq p_{1+}p_{+1}$．我们立刻得到 X 和 Y 不独立．

假设 X 和 Y 的联合概率函数如下表所示．

	Y				
X	1	2	3	4	合计
1	0.06	0.02	0.04	0.08	0.20
2	0.15	0.05	0.10	0.20	0.50
3	0.09	0.03	0.06	0.12	0.30
合计	0.30	0.10	0.20	0.40	1.00

因为表中对于所有 i 和 j(3.5.6)式成立，所以我们得到 X 和 Y 是独立的．——◀

由例3.5.5应注意到 X 和 Y 独立当且仅当他们联合概率函数表中的行相互间成正比，或者等价的，当且仅当表中的列相互间成正比．

现假设 X 和 Y 是有连续型联合分布的随机变量，其概率密度函数为 f. 那么 X 和 Y 是独立的当且仅当对于 $-\infty<x<\infty$ 和 $-\infty<y<\infty$，f 能表示如下： 133

$$f(x,y)=g_1(x)g_2(y). \tag{3.5.7}$$

其中，g_1 是仅与 x 有关的非负函数，g_2 是仅与 y 有关的非负函数. 换言之，它是充分且必要的，对所有的 x 和 y，f 能分解成 x 的任意非负函数和 y 的任意非负函数的积. 然而，需要强调的是，如同(3.5.5)式一般，(3.5.7)式中的因式分解必须对于所有的 x 和 y 成立(即 $-\infty<x<\infty$ 和 $-\infty<y<\infty$).

一个重要的特例是，若取某个边平行于 x 轴和 y 轴的矩形外的所有的 x 和 y 的值时，有 $f(x,y)=0$，则实际上无需对所有的 x 和 y 检验(3.5.7)式. 为了检验本例子中 X 和 Y 是否独立，只需检验对矩形内的所有 x 和 y 的值，(3.5.7)式是否成立. 具体来说，令 a，b，c，d 满足 $-\infty\leqslant a<b\leqslant+\infty$，$-\infty\leqslant c<d\leqslant+\infty$，令其是 xy 平面的下述矩形：

$$S=\{(x,y)\mid a\leqslant x\leqslant b \text{ 且 } c\leqslant y\leqslant d\}. \tag{3.5.8}$$

应注意到任意端点 a，b，c，d 可以是无穷. 假设对于 S 外所有的点 (x,y) 有 $f(x,y)=0$. 那么 X 和 Y 是独立的当且仅当对于所有 S 中的点 f 有如等式(3.5.7)的因式分解.

例 3.5.6 检验联合概率密度函数的因式分解 假设 X 和 Y 的联合概率密度函数如下：

$$f(x,y)=\begin{cases}ke^{-(x+2y)} & \text{对于 } x\geqslant 0 \text{ 且 } y\geqslant 0,\\ 0 & \text{其他.}\end{cases}$$

我们先确定 X 和 Y 是否独立，再确定他们的边际概率密度函数.

本例中，若令等式(3.5.8)中的 $a=0,b=\infty,c=0,d=\infty$，则在 S 外有 $f(x,y)=0$，进一步，在 S 中的每点，令 $g_1=ke^{-x},g_2=e^{-2y}$，则 $f(x,y)$ 有(3.5.7)式的因式分解，因此，X 和 Y 是独立的.

由此可知，本例中，除去常数因子外，$g_1(x)$ 和 $g_2(x)$ 必然是 X 和 Y 当 $x\geqslant 0$ 和 $y\geqslant 0$ 时的边际概率密度函数. 通过选取常数使得 $g_1(x)$ 和 $g_2(y)$ 积分为 1，我们可以得到 X 和 Y 的边际概率密度函数 f_1 和 f_2 必然是：

$$f_1(x)=\begin{cases}e^{-x} & \text{对于 } x\geqslant 0,\\ 0 & \text{其他.}\end{cases}$$

和

$$f_2(y)=\begin{cases}2e^{-2y} & \text{对于 } y\geqslant 0,\\ 0 & \text{其他.}\end{cases}$$

我们将 f_1 乘上 f_2 并比较积与 $f(x,y)$，可以得到 $k=2$. ◀ 134

例 3.5.7 相关随机变量 假设 X 和 Y 的联合概率密度函数如下：

$$f(x,y)=\begin{cases}kx^2y^2 & 对于\ x^2+y^2\leqslant 1,\\ 0 & 其他.\end{cases}$$

我们将说明 X 和 Y 不是独立的.

显然在圆 $x^2+y^2\leqslant 1$ 内的每点，$f(x,y)$ 都有(3.5.7)式的因式分解. 然而，在圆外的每点并不都能满足同样的因式分解. 本例的重要特征是 X 和 Y 的取值限制在圆内，X 和 Y 的联合概率密度函数在圆内为正，在圆外为零. 在此条件下，X 和 Y 不可能独立，因为对于 Y 的所有给定的值 y，X 的可能取值依赖于 y. 比如说，若 $Y=0$，那么 X 能取任意使得 $X^2\leqslant 1$ 的值；又若 $Y=1/2$，那么 X 必须取使得 $X^2\leqslant 3/4$ 的值. ◀

注： 独立随机变量各自的函数是相互独立的. 若 X 和 Y 是独立的，那么 $h(X)$ 和 $g(Y)$ 是独立的，无论 h 和 g 是何种形式. 因为对任意 t，事件 $\{h(X)\leqslant t\}$ 总能写成 $\{X\in A\}$，其中 $A=\{x\mid h(x)\leqslant t\}$. 类似的，$\{g(Y)\leqslant u\}=\{Y\in B\}$，所以由 X 和 Y 满足(3.5.3)式可得 $h(X)$ 和 $g(Y)$ 满足(3.5.4)式.

3.5.3 小结

令 $f(x,y)$ 是随机变量 X 和 Y 的联合概率函数，联合概率密度函数，或联合概率函数/概率密度函数. 记 X 的边际概率函数或概率密度函数为 $f_1(x)$，记 Y 的边际概率函数或概率密度函数为 $f_2(y)$. 为了得到 $f_1(x)$，若 Y 是离散型的，则计算 $\sum_y f(x,y)$，若 Y 是连续型的，则计算 $\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y$. 类似的，为了得到 $f_2(y)$，若 X 是离散型的，则计算 $\sum_x f(x,y)$，若 X 是连续型的，则计算 $\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x$. 随机变量 X 和 Y 是独立的当且仅当对于所有的 x 和 y，$f(x,y)=f_1(x)f_2(y)$. 无论 X 和/或 Y 是离散型的还是连续型的，都是对的. 两个连续型随机变量独立的一个必要条件是所有满足 $f(x,y)>0$ 的点集是矩形. 但这个条件不是充分的.

习题

1. 设 X 和 Y 有连续型联合分布，其联合概率密度函数如下：

$$f(x,y)=\begin{cases}k & 对于\ a\leqslant x\leqslant b\ 且\ c\leqslant y\leqslant d,\\ 0 & 其他.\end{cases}$$

   其中 $a<b$，$c<d$ 且 $k>0$. 求 X 和 Y 的边际分布.

2. 设 X 和 Y 有离散型联合分布，其联合概率函数定义如下：

$$f(x,y)=\begin{cases}\dfrac{1}{30}(x+y) & 对于\ x=0,1,2\ 且\ y=0,1,2,3,\\ 0 & 其他.\end{cases}$$

a. 求 X 和 Y 的边际概率函数.　　b. X 和 Y 是否独立?

3. 设 X 和 Y 有连续型联合分布，其联合概率密度函数如下：

$$f(x,y)=\begin{cases}\frac{3}{2}y^2 & \text{对于 } 0\leqslant x\leqslant 2 \text{ 且 } 0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

a. 求 X 和 Y 的边际概率密度函数.

b. X 和 Y 是否独立?

c. 事件 $\{X<1\}$ 和事件 $\{Y\geqslant 1/2\}$ 是否独立?

135

4. 假设 X 和 Y 的联合分布密度函数如下：

$$f(x,y)=\begin{cases}\frac{15}{4}x^2 & \text{对于 } 0\leqslant y\leqslant 1-x^2,\\ 0 & \text{其他.}\end{cases}$$

a. 确定 X 和 Y 的边际概率密度函数.　　b. X 和 Y 是否独立?

5. 某药店有三个公共电话亭. 令 $p_i(i=0,1,2,3)$ 表示在任意周一晚八点刚好有 i 个电话亭被占用的概率；又假设 $p_0=0.1$，$p_1=0.2$，$p_2=0.4$，$p_3=0.3$. 令 X 和 Y 表示在两个独立的周一晚八点电话亭被占用的个数. 确定：(a) X 和 Y 的联合分布密度函数；(b) $\Pr(X=Y)$；(c) $\Pr(X>Y)$.

6. 假设某药中，某特定化学成分的浓度是有连续分布的随机变量，其概率密度函数 g 如下：

$$g(x)=\begin{cases}\frac{3}{8}x^2 & \text{对于 } 0\leqslant x\leqslant 2,\\ 0 & \text{其他.}\end{cases}$$

假设 X 和 Y 表示两批药中的该化学成分的浓度，它们是独立的且概率密度函数均为 g. 确定：(a) X 和 Y 联合概率密度函数；(b) $\Pr(X=Y)$；(c) $\Pr(X>Y)$；(d) $\Pr(X+Y\leqslant 1)$.

7. 假设 X 和 Y 联合概率密度函数如下：

$$f(x,y)=\begin{cases}2x\mathrm{e}^{-y} & \text{对于 } 0\leqslant x\leqslant 1 \text{ 且 } 0<y<\infty,\\ 0 & \text{其他.}\end{cases}$$

X 和 Y 是否独立?

8. 假设 X 和 Y 联合概率密度函数如下：

$$f(x,y)=\begin{cases}24xy & \text{对于 } x\geqslant 0 \text{ 且 } x+y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

X 和 Y 是否独立?

9. 假设从如下矩形 S 中随机选取点 (X,Y)：

$$S=\{(x,y)\mid 0\leqslant x\leqslant 2 \text{ 且 } 1\leqslant y\leqslant 4\}$$

a. 确定 X 和 Y 联合分布密度函数，X 的边际概率密度函数和 Y 的边际概率密度函数.

b. X 和 Y 是否独立?

10. 假设从如下圆 S 中随机选取点 (X,Y)：

$$S=\{(x,y)\mid x^2+y^2\leqslant 1\}$$

a. 确定 X 和 Y 联合概率密度函数，X 的边际概率密度函数和 Y 的边际概率密度函数.

b. X 和 Y 是否独立?

11. 假设两人要在下午五点至六点间在某地见面，他们约定任何人都不会等对方超过十分钟. 若他们独立而随机的在下午五点至六点间到达，求他们见面的概率.

12. 证明例3.12中的随机变量 X 和 Y 是独立的.（你可能需要求解3.4节的习题9）

3.6　条件分布

我们将条件概率的概念推广到条件分布. 回忆分布事实上就是由随机变量确定的事件的概率的集合. 条件分布就是由依赖于其他随机变量的随机变量确定的事件的概率. 其思想是：在一个应用问题中典型地会有许多我们感兴趣的随机变量. 在我们观测这些随机变量中的一部分后，我们希望能调整那些还未被观测到的随机变量的概率. 给定一个随机变量 Y 的条件下，另一个随机变量 X 的条件分布就是当我们知道 Y 的值后，用来描述 X 的分布.

136

3.6.1　离散型条件分布

在例3.4.2中，我们构造了两个随机变量 X 和 P 的联合概率函数. 那里，X 是表示第一个病人的临床诊断是否成功的指标，P 是一个病人是否成功的概率. (3.4.2)式表示了在给定 $\{P=p\}$ 形式的事件后，$\{X=x\}$ 形式的事件的概率是 $p^x(1-p)^{1-x}$. 那么概率函数对于 $x=0$，1 是 $p^x(1-p)^{1-x}$，作为给定 $P=p$ 后 X 的条件概率函数是有意义的. 事实上，这就是在我们引入一般定义后的情形.

假设 X 和 Y 是两个随机变量，他们有离散型联合概率函数，记其联合概率函数为 f. 如前，我们令 f_1 和 f_2 分别表示 X 和 Y 的边际概率函数. 当观测到随机变量 Y 的值 y 以后，有下述的条件概率给出随机变量 X 取特定值 x 的概率：

$$\Pr(X=x\mid Y=y)=\frac{\Pr(X=x,Y=y)}{\Pr(Y=y)}=\frac{f(x,y)}{f_2(y)}.\tag{3.6.1}$$

换言之，若已知 $Y=y$，那么 X 即是等式(3.6.1)确定的概率的离散分布. 称此分布为给定 $Y=y$ 下 X 的条件分布. 由等式(3.6.1)，对每个使得 $f_2(y)>0$ 的 y，X 的条件分布表示为概率函数 $g_1(x\mid y)$，其定义如下：

$$g_1(x\mid y)=\frac{f(x,y)}{f_2(y)}.\tag{3.6.2}$$

称函数 g_1 为给定 $Y=y$ 下 X 的条件概率函数. 对于每个固定的 y 值，函数 $g_1(x\mid y)$ 是对 X 的所有可能值的概率函数，因为 $g_1(x\mid y)\geqslant 0$ 且

$$\sum_x g_1(x\mid y)=\frac{1}{f_2(y)}\sum_x f(x,y)=\frac{1}{f_2(y)}f_2(y)=1.$$

类似地，若给定 $X=x$，使得 $f_1(x)=\Pr(X=x)>0$，且若 $g_2(y\mid x)$ 是给定 $X=x$ 下

Y 的条件概率函数，那么

$$g_2(y|x)=\frac{f(x,y)}{f_1(x)}. \tag{3.6.3}$$

对于每个确定的 x 值，函数 $g_2(y|x)$ 都是覆盖 Y 的所有可能取值的一个概率函数.

例 3.6.1 由联合概率函数计算条件概率函数 假设 X 和 Y 联合概率函数如例 3.4.1 的表所示. 我们将确定给定 $X=2$ 时 Y 的条件概率函数.

137

由 3.4 节的表，$f_1(2)=\Pr(X=2)=0.6$. 因此，Y 取特定值 y 的条件概率 $g_2(y|2)$ 为

$$g_2(y|2)=\frac{f(2,y)}{0.6}.$$

应该注意的是，对 y 的所有可能值，条件概率 $g_2(y|2)$ 必然正比于联合概率 $f(2,y)$. 本例中，为使得结果值的和等于 1，$f(2,y)$ 的每个值除以常数 $f_1(2)=0.6$. 因此

$$g_2(1|2)=1/2, g_2(2|2)=0, g_2(3|2)=1/6, g_2(4|2)=1/3.$$ ◀

例 3.6.2 临床试验 在讨论离散型条件分布开始时. 我们断言给定 $P=p$，X 的条件分布应该有概率函数 $p^x(1-p)^{1-x}$. 为了确认事实确实如此，利用(3.4.3)式，有 $f(x,p)=p^x(1-p)^{1-x}/11$，对于 $x=0,1$ 及 $p=0,1/10,\cdots,9/10,1$. 现计算

$$f_2(p)=\sum_{x=0}^{1}\frac{1}{11}p^x(1-p)^{1-x}=\frac{1}{11},$$

对于 $p=0,1/10,\cdots,9/10,1$. 由(3.6.2)式，我们得到

$$g_1(x|p)=p^x(1-p)^{1-x}, \text{对于 } x=0,1,$$

它确实是我们预言的概率函数. 我们也能得到对每个 x，给定 $X=x$ 下，P 的条件概率函数. 首先，我们得到

$$f_1(1)=\sum_p \frac{1}{11}p=\sum_{j=1}^{11}\frac{1}{11}\frac{j-1}{10}=\frac{1}{2},$$

这与(2.3.2)式计算得到的结果一致. 同样也有 $f_1(0)=1/2$. 那么，$f(x,p)$ 除以 $f_1(x)$ 得到

$$g_2(p|x)=\frac{p^x(1-p)^{1-x}/11}{f_1(x)}=\frac{2}{11}p^x(1-p)^{1-x}, \text{对于 } p=0,1/10,\cdots,9/10,1.$$

举例来说，如例 2.3.10 的第一部分，若 $x=1$，那么 $g_2(p|1)=\frac{2}{11}p$. 对于 $p=(j-1)/10$，它变成 $g_2([j-1]/10|1)=(j-1)/55$，这与(2.3.8)计算得到的结果一致. 其原因是，如例 3.4.2 中 $\{P=p\}=B_{10p+1}$，$\{X=1\}$是例 2.3.10 中的事件 E_1，那么 $g_2(p|1)=\Pr(B_{10p+1}|E_1)$，这在例(2.3.8)中是计算过的. 综上所述，给定 $X=1$，P 的条件分布与我们所说的，在给定事件 $E_1=\{X=1\}$下，事件$B_1,\cdots,B_{11}$的后验概率是相同的. ◀

138 **注：**随机变量的贝叶斯定理推广的预览. 例3.6.2最后的计算是一般结果的一个例子，它是贝叶斯定理2.3.2的推广. 我们将在处理完连续型和混合型条件分布后给出这个结果的一般形式并证明之.

3.6.2 连续型条件分布

假设 X 和 Y 有连续型联合分布，其联合概率密度函数为 f，边际概率密度函数为 f_1 和 f_2. 假设观测到 $Y=y$，我们想确定 X 取各种可能取值的集合的概率. 应当注意的是，在这种情况下对任意 y 值，$\Pr(Y=y)=0$. 在 $\Pr(B)=0$ 时，$\Pr(A|B)$ 形式的条件概率并没有定义. 因此，为了推导出当 X 和 Y 有连续型联合分布时的条件概率，将通过考察等式(3.6.2)中给出的 X 的条件概率函数以及概率函数与概率密度函数间的相似性来推广条件概率的概念.

令 y 为使 $f_2(y)>0$ 的值. 那么定义给定 $Y=y$，X 的条件概率密度函数的定义如下：

$$g_1(x|y)=\frac{f(x,y)}{f_2(y)} \quad 对 -\infty<x<\infty. \tag{3.6.4}$$

对使得 $f_2(y)=0$ 的那些 y 值，我们可以任意定义 $g_1(x|y)$，然而只希望 $g_1(x|y)$ 作为 x 的函数成为一个概率密度函数. 这样，对于每个固定的 y 值，函数 $g_1(\cdot|y)$ 是 X 在实轴上的概率密度函数. 因为 $g_1(x|y)\geqslant 0$ 且

$$\int_{-\infty}^{\infty} g_1(x|y)\,dx=1.$$

应当注意的是(3.6.2)式和(3.6.4)式是等价的. 然而，(3.6.2)式是通过推导得到的给定 $Y=y$ 时 $X=x$ 的条件分布，而(3.6.4)式是通过定义得到的给定 $Y=y$ 时，X 的条件概率密度函数的值.

可以通过考察图3-18来理解(3.6.4)式给出的定义. 联合概率密度函数 f 定义了 xy 平面上方的一个曲面，它在每个点 (x,y) 的高度 $f(x,y)$ 表示了该点的相对可能性. 例如，若已知 $Y=y_0$，那么点 (x,y) 必然落在 xy 平面的直线 $y=y_0$ 上，且该直线上的任意点 (x,y_0) 的相对可能性就为 $f(x,y_0)$. 因此 X 的条件概率密度函数 $g_1(x|y_0)$ 应与 $f(x,y_0)$ 成比例. 换言之，$g_1(x|y_0)$ 本质上和 $f(x,y_0)$ 相同，不过前者要包括一个常数因子 $1/[f_2(y_0)]$，使得条件概率密度函数对于 x 的所有可能值的积分为1.

类似地，对于任意使 $f_1(x)>0$ 的 x 值，定义给定 $X=x$ 下 Y 的条件概率密度函数如下：

$$g_2(y|x)=\frac{f(x,y)}{f_1(x)} 对 -\infty<y<\infty. \tag{3.6.5}$$

这个等式与对于离散型分布推导出的等式(3.6.3)是一致的. 若 $f_1(x)=0$, $g_2(y|x)$ 可以是任意的，只要 $g_2(y|x)$ 作为 y 的函数成为一个概率密度函数.

f(x,y)
f(x,y₀)
y₀
x
y

图 3-18 条件概率密度函数 $g_1(x|y_0)$ 正比于 $f(x,y_0)$

例 3.6.3 由联合概率密度函数计算条件概率密度函数 假设 X 和 Y 联合概率密度函数如例 3.4.3 中所定义，我们将先确定给定 $X=x$ 下 Y 的条件概率密度函数，再确定给定 $X=1/2$ 时 Y 的一些概率. 139

使得 $f(x,y)>0$ 的点集 S 的图形如图 3-11 所示. 进一步，例 3.5.2 推导的边际概率密度函数 f_1 的图形如图 3-15 所示. 由图 3-15 可见当 $-1<x<1$，且 $x\neq 0$ 时，有 $f_1(x)>0$. 因此，对于每个满足条件的 x 值，Y 的条件概率密度函数 $g_2(y|x)$ 如下：

$$g_2(y|x)=\begin{cases}\dfrac{2y}{1-x^4} & \text{对于 } x^2\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

特别地，若已知 $X=1/2$，那么 $\Pr\left(Y\geqslant\frac{1}{4}\,\middle|\,X=\frac{1}{2}\right)=1$ 且

$$\Pr\left(Y\geqslant\frac{3}{4}\,\middle|\,X=\frac{1}{2}\right)=\int_{3/4}^{1}g_2\left(y\,\middle|\,\frac{1}{2}\right)\mathrm{d}y=\frac{7}{15}.$$

◀

注：条件概率密度函数不是在一个零概率集上取条件得到的结果. 给定 $Y=y$ 条件下 X 的条件概率密度函数 $g_1(x|y)$ 为已知 $Y=y$ 后用以描述 X 的概率密度函数. 看上去 140

似乎是对事件$\{Y=y\}$取条件，而当Y是连续分布时，它是零概率的. 事实上，本文中我们将见到的那些情况，$g_1(x|y)$是极限：

$$g_1(x|y)=\lim_{\varepsilon\to 0}\frac{\partial}{\partial x}\Pr(X\leqslant x|y-\varepsilon<Y\leqslant y+\varepsilon). \tag{3.6.6}$$

若Y的边际概率密度函数在y点是正的，(3.6.6)式中取条件的事件$\{y-\varepsilon\leqslant Y\leqslant y+\varepsilon\}$有正概率. 上式的严格证明所需的数学知识超出了本书的范围. 尽管如此，重要的是需要记住，把给定$Y=y$条件下X的条件概率密度函数看作为给定Y在很靠近y时X的条件概率密度函数，仍是个较好的想法. 这个表述很难使用，我们将不采用，但必须要记住，条件概率密度函数和对一个零概率事件取条件的区别. 除去这些区别，当处理给定$Y=y$条件下X的条件概率密度函数时，把Y当作常数y仍是合理的.

3.6.3　联合分布的构造

在例3.4.2中，X和P的联合概率函数的构造，是条件分布的第二种常用方法. (3.6.2)式和(3.6.3)式表明了如何由联合概率函数和边际概率函数计算条件概率函数. 在例3.4.2中，我们通过条件概率函数和边际概率函数计算了联合概率函数. 当时，我们并没有把$g_1(x|p)=p^x(1-p)^{1-x}$（对$x=0,1$）称为给定$P=p$下X的条件概率函数，因为我们还未引入条件分布. 但在例3.6.2后，我们发现这确实是给定$P=p$下X的条件概率函数. 同样，我们也未把$f_2(p)=1/11$对$p=0,1/10,\cdots,9/10,1$称为P的边际概率函数，但它确实就是. 将它们乘起来给出了$f(x,p)=g_1(x|p)f_2(p)$. 这个构造对一般情形都是可行的，下面将对其进行解释.

1. 对条件概率乘法定理的推广

由(3.6.4)式可知，对每个使得$f_2(y)>0$的值y和每个x值，有

$$f(x,y)=g_1(x|y)f_2(y). \tag{3.6.7}$$

进一步，若对于某个值y_0有$f_2(y_0)=0$，那么，不失一般性，假设对于所有的x值，$f(x,y_0)=0$. 在该种情形下，(3.6.7)式的两边都等于0，而$g_1(x|y_0)$定义的不唯一性变得无关紧要了. 因此，对于所有的x和y的值都满足(3.6.7)式.

类似地，由(3.6.5)式可知，对于所有的x和y的值，$f(x,y)$的联合概率密度函数，能表示如下：

$$f(x,y)=f_1(x)g_2(y|x). \tag{3.6.8}$$

注意到(3.6.7)式和(3.6.8)式与条件概率的乘法原理，即$\Pr(AB)=\Pr(A)\Pr(B|A)$，形式相似. 事实上，对于离散型随机变量的例子，那些等式仅仅是乘法原理的应用. 由于我们对于连续型随机变量的条件概率密度函数的定义，等式(3.6.7)和(3.6.8)对于条件概率密度函数也成立. 事实上，即使f是联合概率函数/概率密度函数，等式(3.6.7)和(3.6.8)也成立.

例 3.6.4 临床试验 在例 3.4.2 中，随机变量 P 是病人被治愈的未知概率，只有取从 0 到 1 的 11 个可能值. P 的一个更实际的分布可能是 $[0,1]$ 整个区间上的连续分布. 假设我们仍令 X 的条件概率函数为 $g_1(x|p)=p^x(1-p)^{1-x}$，对于 $x=0$，1，但令 P 是 $[0,1]$ 区间上的均匀分布. 那么 X 和 P 的联合概率函数/概率密度函数变为：

$$f(x,p)=g_1(x|p)f_2(p)=p^x(1-p)^{1-x},$$

对于 $x=0$，1，$0\leqslant p\leqslant 1$，其他情况为 0. 要得到给定 $X=x$ 下 P 的条件概率密度函数，先确定 X 的边际概率函数：

$$f_1(x)=\int_0^1 p^x(1-p)^{1-x}\mathrm{d}p=\frac{1}{2}, \tag{3.6.9}$$

这与当 P 是离散型分布时得到的结果相同. 那么给定 $X=x$ 下 P 的条件概率密度函数为：

$$g_2(p|x)=\frac{f(x,p)}{f_1(x)}=2p^x(1-p)^{1-x}, \text{对于 } 0<p<1. \tag{3.6.10}$$ ◀

2. 随机变量的贝叶斯定理和全概率原理

(3.6.9) 式中的计算是随机变量全概率原理的一般形式的一个例子. 同样，(3.6.10) 式中的计算，也是随机变量的贝叶斯定理的一般形式的一个例子. 若 $f_2(y)$ 是随机变量 Y 的边际概率函数或概率密度函数，$g_1(x|y)$ 是给定 $Y=y$ 条件下 X 的条件概率函数或概率密度函数，那么联合概率函数，或联合概率密度函数，或联合概率函数/概率密度函数为 $f(x,y)=g_1(x|y)f_2(y)$，X 的边际概率函数或概率密度函数为：

$$f_1(x)=\sum_y g_1(x|y)f_2(y), \tag{3.6.11}$$

其中 Y 是离散型的. 若 Y 是连续型的

$$f_1(x)=\int_{-\infty}^{\infty} g_1(x|y)f_2(y)\mathrm{d}y. \tag{3.6.12}$$

这些等式是随机变量的全概率原理. 为了得到给定 $X=x$ 条件下 Y 的条件分布，将 $f(x,y)$ 除以 $f_1(x)$ 得到：

$$g_2(y|x)=\frac{g_1(x|y)f_2(y)}{f_1(x)}, \tag{3.6.13}$$

即是随机变量的贝叶斯定理. 若联合分布是由 $f_1(x)$ 和 $g_2(y|x)$ 构成的，那么贝叶斯定理为

$$g_1(x|y)=\frac{g_2(y|x)f_1(x)}{f_2(y)}, \tag{3.6.14}$$

其中 $f_2(y)$ 是由全概率原理 (3.6.11) 或 (3.6.12) 中交换 x 与 y 和下标 1 与 2 后得到.

例 3.6.5 从均匀分布中选取点 假设点 X 选自区间 $[0,1]$ 上的均匀分布；当观测到 $X=x$ 后 $(0<x<1)$，由区间 $[x,1]$ 上的均匀分布选取点 Y. 我们将推导出 Y 的边际概率密度函数.

因为 X 是均匀分布，X 的边际概率密度函数如下：

$$f_1(x)=\begin{cases}1 & \text{对于 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

类似地，对于每个 $X=x(0<x<1)$，Y 的条件分布是区间 $[x,1]$ 上的均匀分布. 因为区间长度是 $1-x$，给定 $X=x$ 条件下 Y 的条件概率密度函数是

$$g_2(y|x)=\begin{cases}\dfrac{1}{1-x} & \text{对于 } x<y<1,\\ 0 & \text{其他.}\end{cases}$$

由(3.6.8)式可知，X 和 Y 的联合概率密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{1-x} & \text{对于 } 0<x<y<1,\\ 0 & \text{其他.}\end{cases} \tag{3.6.15}$$

因此，对于 $0<y<1$，Y 的边际概率密度函数 $f_2(y)$ 是

$$f_2(y)=\int_{-\infty}^{\infty}f(x,y)\,\mathrm{d}x=\int_0^y\frac{1}{1-x}\,\mathrm{d}x=-\ln(1-y). \tag{3.6.16}$$

进一步，因为 Y 不可能落在区间 $0<y<1$ 以外，所以，对于 $y\leqslant 0$ 或 $y\geqslant 1$，边际概率密度函数 f_2 如图 3-19 所示. 在该例中函数 f_2 是无界的是值得注意的有趣之处.

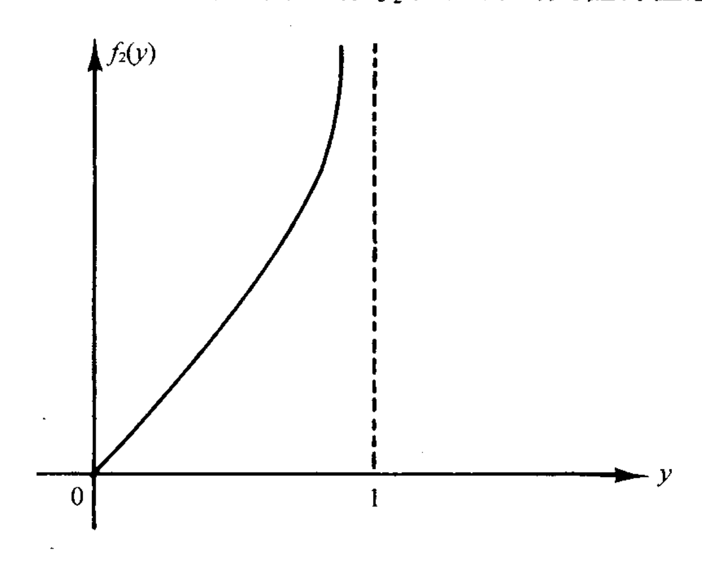


图 3-19 例 3.6.5 中 Y 的边际概率密度函数

利用贝叶斯定理(3.6.14)，我们同样能得到给定 $Y=y$ 条件下 X 的条件概率密度函数. 在(3.6.15)式中已计算过 $f_1(x)$ 和 $g_2(y|x)$ 的乘积. 由(3.6.15)式，它关于

$f_2(y)$的比为

$$g_1(x|y)=\begin{cases}\dfrac{-1}{(1-x)\ln(1-y)} & \text{对于 } 0<x<y,\\ 0 & \text{其他}.\end{cases}$$ ◀

3. 独立随机变量

假设X和Y是两个独立随机变量，它们有连续型联合分布. 由3.5节可知，X和Y是独立的当且仅当它们的联合概率密度函数$f(x,y)$能因式分解为如下形式，对于$-\infty<x<\infty$和$-\infty<y<\infty$：

$$f(x,y)=f_1(x)f_2(y).$$

由(3.6.7)式可知，X和Y是独立的当且仅当对于每个满足$f_2(y)>0$的值y和x的每个值有，

$$g_1(x|y)=f_1(x). \tag{3.6.17}$$

换言之，X和Y是独立的当且仅当对于所有使得$f_2(y)>0$的值y，在给定$Y=y$条件下X的条件概率密度函数和X的边际概率密度函数相等. 因为当$f_2(y)=0$时$g_1(x|y)$是任取的，我们不能期望(3.6.17)式在这种情形下仍成立.

类似地，由(3.6.8)式，X和Y是独立的当且仅当

$$g_2(y|x)=f_2(y). \tag{3.6.18}$$

对于所有使得$f_1(x)>0$的x的每个值均成立.

若两个随机变量X和Y的联合分布是离散型、或是混合型，那么(3.6.17)式和(3.6.18)式成立仍然是X和Y是互相独立的充分必要条件. 然而，在这种情形下，函数f_1、f_2、g_1和g_2对于那些离散型随机变量应被解释为边际和条件概率函数.

注： *条件分布和分布的表现是一样的.* 如同我们在2.1节的注，条件概率和概率的表现是一样的. 因为分布仅仅就是概率的集合，由此可知条件分布和分布的表现是一样的. 举例来说，计算给定$Y=y$离散型随机变量X在区间$[a,b]$的条件概率，我们必须给区间上的每个值x增加$g_1(x|y)$倍. 同样，我们已证明过和将要证明的关于分布的定理都有在附加随机变量上取条件的版本. 我们将这些定理的例子放到3.7节，因为他们依赖于多于两个随机变量的联合分布.

3.6.4 小结

给定另一个随机变量Y的观察值y条件下随机变量X的条件分布，就是在我们知道$Y=y$后，用来描述X的分布. 当处理给定$Y=y$条件下X的条件分布时，把Y作为常数y处理是妥当的. 若X和Y有联合概率函数、概率密度函数或概率函数/概率密度函数$f(x,y)$，那么，给定$Y=y$条件下X的条件概率函数或概率密度函数就是$g_1(x|y)=f(x,y)/f_2(y)$，其中f_2是Y的边际概率函数或概率密度函数. 当直接确定条件分布方便时，能通过条件分布和另一个随机变量的边际分布来构造联合分布.

例如：

$$f(x,y)=g_1(x|y)f_2(y)=f_1(x)g_2(y|x).$$

在这种情形下，我们有全概率原理和贝叶斯定理相应版本，能用来计算另一个的边际和条件分布.

两个随机变量 X 和 Y 是独立的当且仅当对于所有的使得 $f_2(y)>0$ 的值 y，给定 $Y=y$ 条件下 X 的条件概率函数或概率密度函数与 X 的边际概率函数/概率密度函数相等. 等价地，X 和 Y 是独立的当且仅当对于所有的使得 $f_1(x)>0$ 的值 x，给定 $X=x$ 条件下 Y 的条件概率函数或概率密度函数与 Y 的边际概率函数/概率密度函数相等.

习题

1. 假设随机变量 X 和 Y 的联合概率密度函数如例 3.5.7 所示，计算对于每个 y，给定 $Y=y$ 下 X 的条件概率密度函数.
2. 某高校的每个学生按照其在校年数（大一，大二，大三，大四）和其参观某特定博物馆的次数（从未，一次，多于一次）分类. 每个类别中学生的比例如下表：

	从未	一次	多于一次
大一	0.08	0.10	0.04
大二	0.04	0.10	0.04
大三	0.04	0.20	0.09
大四	0.02	0.15	0.10

   a. 若从该高校随机挑选的一个学生是大三的，那么他从未参观过博物馆的概率是多大？
   b. 若从该高校随机挑选的一个学生已经参观过博物馆三次，那么他是大四的概率是多大？
3. 假设从如下圆 S 内随机选择一点（X，Y），S 的定义如下：
$$S=\{(x,y)\mid(x-1)^2+(y+2)^2\leqslant 9\}.$$
确定（a）对于 X 的每个给定的值 Y 的条件概率密度函数，（b）$\Pr(Y>0\mid X=2)$.
4. 假设两个随机变量 X 和 Y 的联合概率密度函数如下：
$$f(x,y)=\begin{cases}c(x+y^2) & 对于\ 0\leqslant x\leqslant 1\ 且\ 0\leqslant y\leqslant 1,\\ 0 & 其他.\end{cases}$$
确定（a）对于每个给定的 Y 的值 X 的条件概率密度函数，（b）$\Pr(X<\frac{1}{2}\mid Y=\frac{1}{2})$.
5. 假设两个随机变量 X 和 Y 的联合概率密度函数如例 3.6.5 中(3.6.15)式所示，确定（a）对于 Y 的每个给定的值，X 的条件概率密度函数，（b）$\Pr(X>\frac{1}{2}\mid Y=\frac{3}{4})$.
6. 假设两个随机变量 X 和 Y 的联合概率密度函数如下：
$$f(x,y)=\begin{cases}c\sin x & 对于\ 0\leqslant x\leqslant \pi/2\ 且\ 0\leqslant y\leqslant 3,\\ 0 & 其他.\end{cases}$$
确定（a）对于 X 的每个给定的值 Y 的条件概率密度函数，（b）$\Pr(1<Y<2\mid X=0.73)$.

7. 假设两个随机变量 X 和 Y 的联合概率密度函数如下：

$$f(x,y)=\begin{cases}\dfrac{3}{16}(4-2x-y) & \text{对于 } x>0,y>0 \text{ 且 } 2x+y<4,\\ 0 & \text{其他.}\end{cases}$$

确定（a）对于 X 的每个给定的值 Y 的条件概率密度函数，（b）$\Pr(Y\geqslant 2\mid X=0.5)$.

145

8. 假设一个人的数学能力测试的分数 X 是 0 到 1 间的一个数，他的音乐能力测试也是 0 到 1 间的一个数. 进一步假设美国的全体大学生中，X 和 Y 有如下的联合概率密度函数：

$$f(x,y)=\begin{cases}\dfrac{2}{5}(2x+3y) & \text{对于 } 0\leqslant x\leqslant 1 \text{ 且 } 0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

a. 大学生在数学测试中获得大于 0.8 分的比例是多少？

b. 如果一个学生的音乐测试分数是 0.3，那么其数学测试分数大于 0.8 的概率有多大？

c. 如果一个学生的数学测试分数是 0.3，那么其音乐测试分数大于 0.8 的概率有多大？

9. 假设能用两种工具来进行某种测量，用工具 1 得到的测量的概率密度函数 h_1 是

$$h_1(x)=\begin{cases}2x & \text{对于 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

用工具 2 得到的测量的概率密度函数 h_2 是

$$h_2(x)=\begin{cases}3x^2 & \text{对于 } 0\leqslant x\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

假设随机选用两种工具中的一种进行测量，并得到结果 X.

a. 确定 X 的边际概率密度函数.

b. 如果测量的值 $X=1/4$，那么工具 1 被选用的概率有多大？

10. 在一堆硬币中，掷不同的硬币得到正面的概率 X 不尽相同，在这堆硬币中 X 的概率密度函数如下：

$$f_1(x)=\begin{cases}6x(1-x) & \text{对于 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

假设从这堆硬币中随机选择一个并掷一次，得到的是正面. 确定这枚硬币的 X 的条件概率密度函数.

11. 当 $f_2(y)=0$ 时，给定 $Y=y$ 条件下 X 的条件概率密度函数的定义是任意的. 但却不造成严重后果的原因是，我们不太可能观测到与使得 $f_2(y_0)=0$ 的 y_0 相接近的 Y. 更严格地说，令 $f_2(y_0)=0$，并令 $A_0=[y_0-\varepsilon,y_0+\varepsilon]$. 同样，令 y_1 满足 $f_2(y_1)>0$，并令 $A_1=[y_1-\varepsilon,y_1+\varepsilon]$. 假设 f_2 在 y_0 点和 y_1 点都连续. 证明：

$$\lim_{\varepsilon\to 0}\frac{\Pr(Y\in A_0)}{\Pr(Y\in A_1)}=0.$$

即，Y 接近于 y_0 的概率比 Y 接近于 y_1 的概率小得多.

12. 令 Y 是交换机收到呼叫的频率（每小时呼叫数）. 令 X 是两小时内呼叫的次数. 假设 Y 的

边际概率密度函数是：

$$f_2(y)=\begin{cases}e^{-y} & 对于 y>0,\\ 0 & 其他.\end{cases}$$

且给定 $Y=y$ 条件下 X 的条件概率密度函数是

$$g_1(x\mid y)=\begin{cases}\dfrac{(2y)^x}{x!}e^{-2y} & 对于 x=0,1,\cdots\\ 0 & 其他.\end{cases}$$

a. 求 X 的边际概率密度函数. (你可以使用公式：$\int_0^\infty y^k e^{-y}\,dy=k!$)

b. 求给定 $X=0$ 条件下 Y 的条件概率密度函数 $g_2(y\mid 0)$.

c. 求给定 $X=1$ 条件下 Y 的条件概率密度函数 $g_2(y\mid 1)$.

d. 求哪些值 y 使得 $g_2(y\mid 1)>g_2(y\mid 0)$？这是否与你呼叫越多、频率越高的直觉相符合？

3.7　多元分布

这一节我们将3.4，3.5和3.6节中的两个随机变量 X 和 Y 的结果推广到任意有限个随机变量 $X_1,\cdots,X_n$. 一般地，称多于两个随机变量的联合分布为多元分布. 统计推断的理论(本书第6章及以后的部分)依赖于观测数据的数学模型，其中每个观测量是随机变量. 为此，多元分布很自然地来源于数据的数学模型中. 其中用得最多的模型是各个随机变量在给定一个或两个其他随机变量下条件独立.

146

3.7.1　联合分布

回忆例2.3.10中的诊断实验. 在实验中，我们观察40个病人，根据实验的结果判断他们是成功还是失败. 对于每个病人，我们观测到只取两个值的随机变量：病人失败记为0，成功记为1. 令 X_i 表示第 i 个病人的随机变量，$i=1,\cdots,40$，令 p 表示一个病人成功的概率. 则我们得到40个我们感兴趣的随机变量.

下面将给出讨论多元分布所需要的重要概念的定义和例子.

1. 联合分布函数

定义 n 个随机变量 $X_1,\cdots,X_n$ 的联合分布函数是 F，它在 n 维空间 $\mathbf{R}^n$ 中每一个点 $(x_1,x_2,\cdots,x_n)$ 的值由以下关系式给出：

$$F(x_1,\cdots,x_n)=\Pr(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n).\qquad(3.7.1)$$

每个多元分布函数与前面给出的一元和二元分布函数有相似的性质.

2. 向量记号

在 n 个随机变量 $X_1,\cdots,X_n$ 的联合分布的学习中，为了方便，通常用向量记号 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 且用 $\boldsymbol{X}$ 表示随机向量. 这样说随机变量 $X_1,\cdots,X_n$ 的联合分布有联合分布函数 $F(x_1,\cdots,x_n)$，可以简述为随机向量 $\boldsymbol{X}$ 的联合分布有分布函数 $F(\boldsymbol{x})$. 当使

用这个向量记号时，必须记住 $\boldsymbol{X}$ 是 n 维随机向量，则定义它的分布函数为 n 维空间 $\mathbf{R}^n$ 上的函数. 在任一点 $\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n$,$F(\boldsymbol{x})$的值由(3.7.1)式所确定.

3. 离散分布

称 n 个随机变量 $X_1,\cdots,X_n$ 具有*离散的联合分布*，如果随机向量$(X_1,\cdots,X_n)$只取 R^n 中有限个或可数个不同的值$(x_1,x_2,\cdots,x_n)$. 此时定义 $X_1,\cdots,X_n$ 的联合概率函数为 R^n 中的函数f, 使得对任意的$(x_1,\cdots,x_n)\in\mathbf{R}^n$

$$f(x_1,\cdots,x_n)=\Pr(X_1=x_1,\cdots,X_n=x_n).$$

若用向量记号，则称随机向量 $\boldsymbol{X}$ 具有离散分布，且它的概率函数在每一点$\boldsymbol{x}\in\mathbf{R}^n$ 由以下关系式所确定

$$f(\boldsymbol{x})=\Pr(\boldsymbol{X}=\boldsymbol{x}).$$

147

对于任意的 $A\subset\mathbf{R}^n$,

$$\Pr(\boldsymbol{X}\in A)=\sum_{\boldsymbol{x}\in A}f(\boldsymbol{x})$$

容易证明，如果每个 $X_1,\cdots,X_n$ 都具有离散型分布，则 $\boldsymbol{X}=(X_1,\cdots,X_n)$具有离散型联合分布.

4. 连续分布

称 n 个随机变量 $X_1,\cdots,X_n$ 具有连续型联合分布，如果存在定义在 $\mathbf{R}^n$ 上的非负函数f, 使得对任意的子集 $A\subset\mathbf{R}^n$,

$$\Pr[(X_1,\cdots,X_n)\in A]=\int\underset{A}{\cdots}\int f(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n. \qquad (3.7.2)$$

称函数f为 $X_1,\cdots,X_n$ 的*联合概率密度函数*.

如果 $X_1,\cdots,X_n$ 的联合分布是连续型，则概率密度函数f可由以下关系从联合分布函数 F 导出：

$$f(x_1,\cdots,x_n)=\frac{\partial^n F(x_1,\cdots,x_n)}{\partial x_1\cdots\partial x_n}$$

其中上式中的导数在所有$(x_1,\cdots,x_n)$点存在.

若用向量记号，$f(\boldsymbol{x})$表示随机向量 $\boldsymbol{X}$ 的概率密度函数，(3.7.2)式可以简写为：

$$\Pr(\boldsymbol{X}\in A)=\int\underset{A}{\cdots}\int f(x)\,\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

需要注意，即使每个 $X_1,\cdots,X_n$ 都具有连续分布，向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$也不一定具有连续分布. 见该节习题9.

例3.7.1 排队中的服务时间 排队是指一些顾客为得到服务成排队并按某种算法接受服务. 一个简单的模型是只有一个服务员按照顾客排队的顺序为他们服务. 假设 n 个顾客来到只有单个服务员队伍接受服务. 设 X_i 表示服务员为第 i 个顾客服务的时间，$i=1,\cdots,n$. 我们可以用 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 下面形式的联合密度函数 $f(\boldsymbol{x})$. 对此有

$$f(\boldsymbol{x})=\begin{cases}\dfrac{c}{(2+\sum_{i=1}^{n}x_i)^{n+1}} & 对于\ x_i>0,\\ 0 & 其他.\end{cases} \tag{3.7.3}$$

现在来确定 c 使得(3.7.3)式中的函数为联合概率密度函数. 可以通过对于每一个变量 $x_1,\cdots,x_n$ 作积分(从 x_n 开始)来做到这一点. 第一步积分是

148

$$\int_0^\infty \frac{c}{(2+x_1+\cdots+x_n)^{n+1}}\mathrm{d}x_n=\frac{c/n}{(2+x_1+\cdots+x_{n-1})^n}. \tag{3.7.4}$$

方程(3.7.4)的右端与原概率密度函数的形式是一样的，除了 n 换成了 $n-1$，c 换成了 c/n. 因此对变量 x_i 积分($i=n-1,n-2,\cdots,1$)，结果有相同形式只是 n 换成 $i-1$且 c 换成 $c/n(n-1)\cdots i$. 对于 $\boldsymbol{x}$ 的所有分量积分得到的最后结果是$c/[2(n!)]$. 令它等于1，可得 $c=2(n!)$. ◀

5. 混合型分布

在例3.6.4疾病诊断例子中，其中一个随机变量（即没复发）具有连续型分布，其他的（对于各个病人的成功的示性函数）具有离散型分布. 一般地，如果 $X_1,\cdots,X_n$ 是随机变量，其中一些具有连续型分布，另一些具有离散型分布，则它们的联合分布可以用我们称之为联合概率函数/概率密度函数的函数来表示. 可以通过对于一些分量的值求和且对于另一些分量的值求积分得到$(X_1,\cdots,X_n)\in A\subset\mathbf{R}^n$ 的概率.

例3.7.2　到达一个排队时刻　在例3.7.1中我们介绍了有单个服务员的排队并讨论了服务时间. 影响排队的运作的另一个因素是到达分布，即两个顾客到达之间的时间的分布或者在给定时间内到达的顾客的人数的分布. 如果用 Y 表示顾客到达的速率，Z 表示一个小时内到达的顾客的人数，则典型地可以设 Y 具有连续型分布，Z 具有离散型分布. ◀

3.7.2　边际分布

1. 边际概率密度函数的推导

如果已知 n 个随机变量 $X_1,\cdots,X_n$ 的联合分布. 则可由联合分布导出每个随机变量 X_i 的边际分布. 例如，如果 $X_1,\cdots,X_n$ 的联合概率密度函数是f，则 X_1 的边际概率密度函数f_1 在每个值 x_1 处可由以下关系式确定：

$$f_1(x_1)=\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-1}f(x_1,\cdots,x_n)\,\mathrm{d}x_2\cdots\mathrm{d}x_n.$$

更一般地，可以通过对其他 $n-k$ 个变量积分得到 n 个随机变量 $X_1,\cdots,X_n$ 中任意 k 个的边际概率密度函数. 比如设f是4个随机变量 X_1，X_2，X_3，X_4 联合概率密度函数，则 X_2，X_4 的二元边际概率密度函数f_{24}在任一点(x_2,x_4)的值为

$$f_{24}(x_2,x_4)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x_1,x_2,x_3,x_4)\,\mathrm{d}x_1\mathrm{d}x_3.$$

例 3.7.3 排队中的服务时间 设在例 3.7.1 中 $n=5$，我们来求(X_1,X_4)的二元边际概率密度函数. 必须将(3.7.3)式对 x_2，x_3 和 x_5 积分. 因为联合概率密度函数关于 $\boldsymbol{x}$ 的分量的置换对称，我们只需对后 3 个变量积分，然后将剩下的变量换成 x_1 和 x_4. 在例 3.7.1 已经看到其过程. 结果是

149

$$f_{12}(x_1,x_2)=\begin{cases}\dfrac{4}{(2+x_1+x_2)^3} & \text{对于 } x_1,x_2>0,\\ 0 & \text{其他.}\end{cases} \tag{3.7.5}$$

则 f_{14} 就是将(3.7.5)式中的下标 2 换成 4. 每个 x_i 的一元边际概率密度函数为

$$f_i(x_i)=\begin{cases}\dfrac{2}{(2+x_i)^2} & \text{对于 } x_i>0,\\ 0 & \text{其他.}\end{cases} \tag{3.7.6}$$

所以，例如，如果我们想知道一个顾客等待的时间超过 3 个单位的概率，可以通过对(3.7.6)式中的函数从 3 到 ∞ 积分来计算 $P_i(X_i>3)$，结果是 0.4. ◄

如果 n 个随机变量 $X_1,\cdots,X_n$ 有离散的联合分布，则可以用类似于连续型分布的关系求出 n 个变量中的任意子集的边际联合概率函数. 只是将积分换成求和.

2. 边际分布函数的推导

现在考虑联合分布函数是 F 的 $X_1,\cdots,X_n$ 的联合分布. 可由以下关系式得到 X_1 的边际分布函数 F_1:

$$F_1(x_1)=\Pr(X_1\leqslant x_1)=\Pr(X_1\leqslant x_1,X_2<\infty,\cdots,X_n<\infty=\lim_{x_2,\cdots,x_n\to\infty}F(x_1,x_2,\cdots,x_n).$$

更一般地，n 个随机变量 $X_1,\cdots,X_n$ 中任意 k 个的边际联合分布函数可以通过计算其他 $n-k$ 个变量 x_j 当 $x_j\to\infty$ 时 F 的极限得到. 例如，如果 F 是四个随机变量 X_1，X_2，X_3 和 X_4 的联合分布函数，则 X_2 和 X_4 的二维边际分布函数 F_{24} 在任意一点(x_2,x_4)的值为:

$$F_{24}(x_2,x_4)=\lim_{x_1,x_3\to\infty}F(x_1,x_2,x_3,x_4).$$

3. 独立随机变量

称 n 个随机变量 $X_1,\cdots,X_n$ 是**独立的**，若对于任意 n 个实数集 $A_1,A_2,\cdots,A_n$ 有:

$$\begin{aligned}&\Pr(X_1\in A_1,X_2\in A_2,\cdots,X_n\in A_n)\\&=\Pr(X_1\in A_1)\Pr(X_2\in A_2)\cdots\Pr(X_n\in A_n).\end{aligned}$$

如果 $X_1,\cdots,X_n$ 是独立的，容易得到随机变量 $X_1,\cdots,X_n$ 的任意非空子集也是独立的(见习题 11). 如果令 F 表示 $X_1,\cdots,X_n$ 的联合分布函数，F_i 表示 X_i 的一维边际分布函数，$i=1,\cdots,n$,则由独立性的定义知道 $X_1,\cdots,X_n$ 独立的充分必要条件是，对于任意的$(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$,

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n).$$

换言之,随机变量$X_1,\cdots,X_n$独立的充分必要条件是:如果它们的联合分布函数

150　是 n 个一维边际分布函数的乘积.

此外，如果变量 $X_1,\cdots,X_n$ 具有连续型、离散型或混合型联合分布，使得相应的联合概率密度函数、联合函数或联合概率函数/概率密度函数是 f，且 f_i 是单个变量 X_i 的边际一元概率密度函数或概率函数，$i=1,\cdots,n$ 则 $X_1,\cdots,X_n$ 独立的充分必要条件是对于所有的点 $(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$ 有以下关系式：

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)f_2(x_2)\cdots f_n(x_n). \tag{3.7.7}$$

例 3.7.4　排队中的服务时间　在例 3.7.3 中，我们可以利用(3.7.6)计算 X_1 和 X_2 的两个一元边际概率密度函数的相乘，且可以看出这一乘积并不等于(3.7.5)式中 (X_1,X_2) 的二元边际概率密度函数. 所以 X_1 和 X_2 不是独立的. ◀

4. 随机样本

考虑实数轴上一个给定的概率分布，它可以用一个概率函数或概率密度函数 f 表示. 称 n 个随机变量 $X_1,\cdots,X_n$ 是该分布的一个随机样本，如果这些变量是独立的且边际概率函数或边际概率密度函数是 f. 换言之，即变量 $X_1,\cdots,X_n$ 组成由 f 表示的分布的一个随机样本，如果它们的联合概率函数或联合概率密度函数 g 由下式确定，对任意的 $(x_1,x_2,\cdots,x_n)\in\mathbf{R}^n$ 有：

$$g(x_1,x_2,\cdots,x_n)=f_1(x_1)f_2(x_2)\cdots f_n(x_n).$$

因此，随机样本中的变量是独立同分布的，通常简记为 i. i. d.

若用该形式，说 $X_1,\cdots,X_n$ 组成概率密度函数为 $f(x)$ 的分布的随机样本等价于 $X_1,\cdots,X_n$ 是独立同分布的，且有共同的概率密度函数 $f(x)$. 称 n 为样本容量.

例 3.7.5　灯泡的寿命　假设一个工厂生产的灯泡的寿命服从下列概率密度函数的分布：

$$f(x)=\begin{cases}xe^{-x} & \text{对于 } x>0,\\ 0 & \text{其他.}\end{cases}$$

下面来推导从工厂的产品中随机抽取 n 个灯泡的寿命的联合概率密度函数.

抽取的灯泡的寿命 $X_1,\cdots,X_n$ 组成概率密度函数为 f 的随机样本. 来了方便，当 v 的表达式复杂时，我们用记号 $\exp(v)$ 表示指数 e^v. 则 $X_1,\cdots,X_n$ 的联合概率密度函数 g 由下式给出，当 $x_i>0$，$i=1$，$\cdots n$ 时

$$g(x_1,\cdots,x_n)=\prod_{i}^{n}f(x_i)=\left(\prod_{i=1}^{n}x_i\right)\exp\left(-\prod_{i=1}^{n}x_i\right).$$

对于其他点，$g(x_1,\cdots,x_n)=0$.

151　涉及 n 个寿命 $X_1,\cdots,X_n$ 的每个概率原则上可以通过在 $\mathbf{R}^n$ 的恰当的子集上的积分得到. 比如，如果 A 中的点 $(x_1,\cdots,x_n)$ 满足 $x_i>0,i=1,\cdots,n$. 且 $\sum_{i=1}^{n}x_i<a$，其中

a 是给定的正数，则

$$\Pr\left(\sum_{i=1}^{n} x_i < a\right) = \int \underset{A}{\cdots} \int \left(\prod_{i=1}^{n} x_i\right) \exp\left(-\sum_{i=1}^{n} x_i\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n. \quad \blacktriangleleft$$

如果不借助于积分表或计算机. 例3.7.5后面的给定的积分的计算将花费大量的时间. 然而，可以利用连续分布和随机样本的性质简化一些特定的概率的计算. 例如，已知例3.7.5的条件，要计算 $\Pr(X_1 < X_2 < \cdots < X_n)$. 因为变量 $X_1, \cdots, X_n$ 具有连续型联合分布，至少有两个随机变量取相同值的概率为0. 事实上，向量 $X_1, \cdots, X_n$ 属于 $\mathbf{R}^n$ 的某个其 n 维体积为0的子集的概率为0. 而且，因为 $X_1, \cdots, X_n$ 是独立同分布的，每一个变量都等概地可能成为 n 个寿命中最小值，每一个也等概地可能成为最大值. 更一般地，如果寿命 $X_1, \cdots, X_n$ 按照由小到大排列，每一个特定的顺序都等概地可能成为另一个顺序. 因为有 $n!$ 个不同的顺序，因此每一个特定顺序 $X_1 < X_2 < \cdots < X_n$ 的概率是 $1/n!$. 从而，

$$\Pr(X_1 < X_2 < \cdots < X_n) = \frac{1}{n!}.$$

3.7.3 条件分布

假设 n 个随机变量 $X_1, \cdots, X_n$ 的联合分布是连续型的，其联合概率密度函数是 f，用 f_0 表示 $n-1$ 个变量 $X_2, \cdots, X_n$ 的边际联合概率密度函数. 则对于所有使 $f_0(x_2, \cdots, x_n) > 0$ 的 $x_2, \cdots, x_n$，在给定 $X_2 = x_2, \cdots, X_n = x_n$ 的条件下，X_1 的条件概率密度函数为：

$$g_1(x_1 \mid x_2, \cdots, x_n) = \frac{f(x_1, x_2, \cdots, x_n)}{f_0(x_2, \cdots, x_n)}.$$

更一般地，假设随机向量 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 被分成两个子向量 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$，$\boldsymbol{Y}$ 是由组成 $\boldsymbol{X}$ 的 n 个随机变量中的 k 个组成的 k 维随机向量，$\boldsymbol{Z}$ 是剩下的 $n-k$ 个随机变量组成的 $n-k$ 维随机向量. 同时假设 n 维向量 $(\boldsymbol{Y}, \boldsymbol{Z})$ 的概率密度函数是 f，$\boldsymbol{Z}$ 的 $n-k$ 维边际概率密度函数是 f_2. 则对于每个满足 $f_2(\boldsymbol{z}) > 0$ 的点，$\boldsymbol{z} \in \mathbf{R}^{n-k}$，定义当 $\boldsymbol{Z} = \boldsymbol{z}$ 时，$\boldsymbol{Y}$ 的 k 维条件概率密度函数 g_1 为：

$$g_1(\boldsymbol{y} \mid \boldsymbol{z}) = \frac{f(\boldsymbol{y}, \boldsymbol{z})}{f_2(\boldsymbol{z})} \quad \text{对于 } \boldsymbol{y} \in \mathbf{R}^k. \tag{3.7.8}$$

也可将方程(3.7.8)写为

$$f(\boldsymbol{y}, \boldsymbol{z}) = g_1(\boldsymbol{y} \mid \boldsymbol{z}) f_2(\boldsymbol{z}), \tag{3.7.9}$$

152

该式表明可以由条件概率密度函数和边际概率密度函数来求联合概率密度函数. 像一元的情形，可以假设如果 $f_2(\boldsymbol{z}) = 0, f(\boldsymbol{y}, \boldsymbol{z}) = 0$. 则(3.7.9)式对于任意的 $\boldsymbol{y}$ 和 $\boldsymbol{z}$ 成立，即使 $g_1(\boldsymbol{y} \mid \boldsymbol{z})$ 不是唯一定义的.

例3.7.6 排队中的服务时间 在例3.7.3中，我们计算了两个服务时间 $\boldsymbol{Z}=(X_1, X_2)$ 的边际二元分布. 现在我们对于任意满足 x_1，$x_2>0$ 的 (x_1,x_2)，可以求在给定 $\boldsymbol{Z}=(X_1,X_2)$ 条件下 $\boldsymbol{Y}=(X_3,X_4,X_5)$ 的条件3维概率密度函数：

$$
\begin{aligned}
g_1(x_3,x_4,x_5\mid x_1,x_2) &= \frac{f(x_1,\cdots,x_5)}{f_{12}(x_1,x_2)} \\
&= \left(\frac{240}{(2+x_1+\cdots+x_5)^6}\right)\left(\frac{4}{(2+x_1+x_2)^3}\right)^{-1} \\
&= \frac{60(2+x_1+x_2)^3}{(2+x_1+\cdots+x_5)^6},\ x_3,x_4,x_5>0
\end{aligned}
$$

对其他点，$g_1(x_3,x_4,x_5\mid x_1,x_2)=0$. ◀

如果随机向量 $\boldsymbol{X}=(\boldsymbol{Y},\boldsymbol{Z})$ 具有离散型或混合型的联合分布，其联合概率函数或概率函数/概率密度函数是 f，且如果 $\boldsymbol{Z}$ 的边际概率函数或概率函数/概率密度函数是 f_2，则对给定 $\boldsymbol{Z}=\boldsymbol{z}$ 时，$\boldsymbol{Y}$ 的条件概率函数或概率函数/概率密度函数 $g_1(\boldsymbol{y}\mid\boldsymbol{z})$ 由(3.7.8)式给出. 且(3.7.9)式也成立.

例3.7.7 确定边际二元概率密度函数 假设 Z 是随机变量有如下概率密度函数 f_0：

$$
f_0(z)=\begin{cases}2\mathrm{e}^{-2z} & \text{对于 } z>0,\\ 0 & \text{其他.}\end{cases} \tag{3.7.10}
$$

并且，假设对于任意给定的 $Z=z>0$，其他的两个随机变量 X_1 和 X_2 是独立同分布的，每一个的条件概率密度函数是：

$$
g(x\mid z)=\begin{cases}z\mathrm{e}^{-zx} & \text{对于 } x>0,\\ 0 & \text{其他.}\end{cases} \tag{3.7.11}
$$

我们来确定 (X_1,X_2) 的边际联合概率密度函数.

因为 X_1 和 X_2 对于给定的 Z 值是 i.i.d. 所以当 $Z=z>0$ 时，它们的条件联合概率密度函数是：

$$
g_{12}(x_1,x_2\mid z)=\begin{cases}z^2\mathrm{e}^{-z(x_1+x_2)} & x_1,x_2>0,\\ 0 & \text{其他.}\end{cases}
$$

(Z,X_1,X_2) 的联合概率密度函数 f 仅在满足 x_1，x_2，$z>0$ 的点 (z,x_1,x_2) 是正的. 因此，在每个这样的点有

153

$$
f(z,x_1,x_2)=f_0(z)g_{12}(x_1,x_2\mid z)=2z^2\mathrm{e}^{-z(2+x_1+x_2)}.
$$

对于 $x_1>0$ 和 $x_2>0$，可以通过分步积分或5.9节中将得到的某些特殊结果来计算 X_1 和 X_2 的边际联合概率密度函数 $f_{12}(x_1,x_2)$：

$$
f_{12}(x_1,x_2)=\int_0^\infty f(z,x_1,x_2)\,\mathrm{d}z=\frac{4}{(2+x_1+x_2)^3}
$$

对 x_1, $x_2>0$. 可以看出这个概率密度函数与(3.7.5)式得出的(X_1,X_2)的边际二元概率密度函数一致.

由这个边际二元概率密度函数可以计算涉及 X_1 和 X_2 的概率，诸如 $\Pr(X_1+X_2<4)$. 我们有

$$\Pr(X_1+X_2<4)=\int_0^4\int_0^{4-x_2}\frac{4}{(2+x_1+x_2)^3}\mathrm{d}x_1\mathrm{d}x_2=\frac{2}{9}.$$

下面来计算给定 $X_1=x_1$, $X_2=x_2$, 对于任意 x_1, $x_2>0$ 下，Z 的条件概率密度函数. 对于任意的 z, 条件概率密度函数是：

$$g_0(z\mid x_1,x_2)=\frac{f(z,x_1,x_2)}{f_{12}(x_1,x_2)}$$

$$=\begin{cases}\frac{1}{2}(2+x_1+x_2)^3z^2\mathrm{e}^{-z(2+x_1+x_2)} & \text{对于 } z>0,\\ 0 & \text{其他.}\end{cases}\tag{3.7.12}$$

最后，我们计算 $\Pr(Z\leqslant 1\mid X_1=1,X_2=4)$.

$$\Pr(Z\leqslant 1\mid X_1=1,X_2=4)=\int_0^1 g_0(z\mid 1,4)\mathrm{d}z$$

$$=\int_0^1 171.5z^2\mathrm{e}^{-7z}\,\mathrm{d}z=0.9704.$$ ◀

1. 全概率公式和贝叶斯定理

例3.7.7包含了多元形式的全概率公式和贝叶斯定理，下面将给出一般性的证明. 我们将利用例(3.7.8)和(3.7.9)中的记号. $\boldsymbol{Y}$ 的边际概率密度函数是：

$$f_1(\boldsymbol{y})=\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-k}g_1(\boldsymbol{y}\mid\boldsymbol{z})f_2(\boldsymbol{z})\mathrm{d}z_1\cdots\mathrm{d}z_{n-k}\tag{3.7.13}$$

给定 $\boldsymbol{Y}=\boldsymbol{y}$ 下 $\boldsymbol{Z}$ 的条件概率密度函数是：

$$g_2(\boldsymbol{z}\mid\boldsymbol{y})=\frac{g_1(\boldsymbol{y}\mid\boldsymbol{z})f_2(\boldsymbol{z})}{f_1(\boldsymbol{y})}\tag{3.7.14}$$

(3.7.14)式是任意多元分布的贝叶斯定理，包括离散型，连续型和混合型. (3.7.13)式是多元分布的全概率公式，其中 $\boldsymbol{Z}$ 具有连续的联合分布. 如果 $\boldsymbol{Z}$ 具有离散型联合分布，(3.7.13)式中的多重积分需要换成多重求和. 如果 $\boldsymbol{Z}$ 具有混合型的联合分布，则多重积分需要换成对连续部分积分和对离散部分求和. (在例3.7.7中，$\boldsymbol{Z}$ 是一元随机变量 Z 和 $\boldsymbol{Y}=(X_1,X_2)$.)

154

例3.7.7也包含了利用随机变量条件独立性的例子. 即给定 $Z=z$, $z>0$ 时，X_1 和 X_2 是条件独立的. 考虑 X_1 和 X_2 是为排队中顾客的服务时间，在5.9节中，我们将看到如何将 Z 解释为顾客被服务的速率. 如果该速率是未知的，则成为不确定性的主要原因. 根据速率 Z 的值划分样本空间然后限制在每一个值上可以消除不确定性的一个主要来源.

一般地，随机变量的条件独立性类似于事件的条件独立性. $\boldsymbol{Z}$ 表示联合概率函数、概率密度函数或概率函数/概率密度函数为 $f_0(\boldsymbol{z})$ 的随机向量. 对于满足 $f_0(\boldsymbol{z})>0$ 的所有 $\boldsymbol{z}$，随机变量 X_1，…，X_n 是给定 $\boldsymbol{Z}$ 下条件独立的，我们有，

$$g(\boldsymbol{x}|\boldsymbol{z}) = \prod_{i=1}^{n} g_i(x_i|\boldsymbol{z}),$$

$g(\boldsymbol{x}|\boldsymbol{z})$ 表示给定 $\boldsymbol{Z}=\boldsymbol{z}$ 时 $\boldsymbol{X}$ 的条件多元概率函数、概率密度函数或概率函数/概率密度函数，$g_i(x_i|\boldsymbol{z})$ 表示给定 $\boldsymbol{Z}=\boldsymbol{z}$ 时 X_i 的一元条件概率函数或概率密度函数. 在例3.7.7中 $g_i(x_i|z)=ze^{-zx_i}$，$x_i>0$.

2. 过去和将来定理的条件形式

我们前面已经说明条件分布和分布的性质类似. 因此，我们已经和将要证明的定理都有相应的条件形式. 例如，(3.7.13)式中的全概率公式、定律有以下的在对另一个随机向量 $\boldsymbol{W}=\boldsymbol{w}$ 下的条件形式：

$$f_1(\boldsymbol{y}|\boldsymbol{w}) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-k} g_1(\boldsymbol{y}|\boldsymbol{z},\boldsymbol{w})f_2(\boldsymbol{z}|\boldsymbol{w})\,\mathrm{d}z_1\cdots\mathrm{d}z_{n-k}, \tag{3.7.15}$$

其中 $f_1(\cdot|\boldsymbol{w})$ 表示给定 $\boldsymbol{W}=\boldsymbol{w}$ 时 $\boldsymbol{Y}$ 的条件概率密度函数、概率函数或概率函数/概率密度函数，$g_1(\cdot|\boldsymbol{z},\boldsymbol{w})$ 表示给定 $(\boldsymbol{Z},\boldsymbol{W})=(\boldsymbol{z},\boldsymbol{w})$ 时 $\boldsymbol{Y}$ 的条件概率密度函数、概率函数或概率函数/概率密度函数，$f_2(\cdot|\boldsymbol{w})$ 表示给定 $\boldsymbol{W}=\boldsymbol{w}$ 时 $\boldsymbol{Z}$ 的条件概率密度函数. 使用同样的记号，贝叶斯定理的条件形式是：

$$g_2(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{w}) = \frac{g_1(\boldsymbol{y}|\boldsymbol{z},\boldsymbol{w})f_2(\boldsymbol{z}|\boldsymbol{w})}{f_1(\boldsymbol{y}|\boldsymbol{w})}. \tag{3.7.16}$$

例3.7.8　基于序列中的随机变量的条件　在例3.7.7中，我们得到了给定 $(X_1,X_2)=(x_1,x_2)$ 时 Z 的条件概率密度函数. 现在假设增加了三个观测变量 X_3、X_4 和 X_5，并且假设给定 $Z=z$ 时 $X_1,\cdots,X_5$ 是i. i. d，其概率密度函数为 $g(x|z)$. 我们使用贝叶斯定理的条件形式来计算给定 $(X_1,\cdots,X_5)=(x_1,\cdots,x_5)$ 时 Z 的条件概率密度函数. 首先，我们来求在 $Z=z$ 和 $\boldsymbol{W}=(X_1,X_2)=(x_1,x_2)$ 的条件下 $\boldsymbol{Y}=(X_3,X_4,X_5)$ 的条件概率密度函数 $g_{345}(x_3,x_4,x_5|x_1,x_2,z)$. 我们将使用这个例子前面的概率密度函数的记号. 因为 $X_1,\cdots,X_5$ 是给定 Z 下条件i. i. d，就有 $g_1(\boldsymbol{y}|z,\boldsymbol{w})$ 不依赖于 $\boldsymbol{w}$. 事实上

155

$$g_1(\boldsymbol{y}|z,\boldsymbol{w}) = g(x_3|z)g(x_4|z)g(x_5|z) = z^3e^{-z(x_3+x_4+x_5)},\quad x_3,x_4,x_5>0$$

根据(3.7.12)式计算得到的在 $\boldsymbol{W}=\boldsymbol{w}$ 的条件下 Z 的条件概率密度函数，记为 $f_2(z|\boldsymbol{w})$，

$$f_2(z|\boldsymbol{w}) = \frac{1}{2}(2+x_1+x_2)^3z^2e^{-z(2+x_1+x_2)}.$$

最后，根据例3.7.6的计算，我们需要在已知前两个观测量的条件下，后三个观测量的条件概率密度函数为：

$$f_1(\boldsymbol{y}|\boldsymbol{w}) = \frac{60(2+x_1+x_2)^3}{(2+x_1+\cdots+x_5)^6}.$$

现在利用贝叶斯定理(3.7.16)结合这些关系式得到

$$g_2(z|\boldsymbol{y},\boldsymbol{w})=\frac{z^3\mathrm{e}^{-z(x_3+x_4+x_5)}\frac{1}{2}(2+x_1+x_2)^3z^2\mathrm{e}^{-z(2+x_1+x_2)})}{\frac{60(2+x_1+x_2)^3}{(2+x_1+\cdots+x_5)^6}}$$

$$=\frac{1}{120}(2+x_1+\cdots+x_5)^6z^5\mathrm{e}^{-z(2+x_1+\cdots+x_5)},$$

对于 $z>0$ 成立. ◀

注：创建结果的条件形式的一些简单法则. 若想确定已经证明的结果在 $\boldsymbol{W}=\boldsymbol{w}$ 的条件下的一个条件形式. 只要将“条件 $\boldsymbol{W}=\boldsymbol{w}$”加到结果的每一个概率的表达式中. 包括所有的概率、分布函数、分位数、分布的名称、概率密度函数、密度函数等等. 也包括以后章节中要介绍的一些与概率有关的概念. (例如第4章中介绍的期望值和方差.)

3.7.4 小结

称随机变量的一个有限集为随机向量. 我们已经定义了任意随机向量的联合分布. 每一个随机向量都有联合分布函数. 连续型随机向量有连续的联合密度函数，离散型随机向量具有离散的联合概率函数，混合型随机向量具有联合的概率函数/概率密度函数. 称 n 维随机向量 $\boldsymbol{X}$ 的分量互相独立的，如果联合概率函数、概率密度函数或概率函数/概率密度函数 $f(\boldsymbol{x})$ 可以分解为，$f(\boldsymbol{x})=\prod_{i=1}^{n}f_i(x_i)$.

我们可以计算随机向量的子向量的边际分布，也可以计算已知一个向量的补向量的条件下它的条件分布. 也可以根据一个随机向量的部分向量的边际分布以及它的补向量的条件分布来计算联合分布. 对于随机向量，也有相应的贝叶斯定理和全概率公式.

一个 n 维随机向量 $\boldsymbol{X}$ 的分量在给定的 $\boldsymbol{Z}$ 下是条件独立的，如果在 $\boldsymbol{Z}=\boldsymbol{z}$ 的条件下 $\boldsymbol{X}$ 的条件概率函数、概率密度函数或概率函数/概率密度函数 $g(\boldsymbol{x}|\boldsymbol{z})$ 可以分解为 $\prod_{i=1}^{n}g_i(x_i|\boldsymbol{z})$. 对于贝叶斯定理和全概率公式以及将来要学到的关于随机变量和随机向量的定理，都存在任意一个附加随机向量 $\boldsymbol{W}$ 下的条件形式.

156

习题

1. 设三个随机变量 X_1，X_2 和 X_3 有连续型的联合分布，其联合概率密度函数为：

$$f(x_1,x_2,x_3)=\begin{cases}c(x_1+2x_2+3X_3) & 0\leqslant x_i\leqslant 1(i=1,2,3),\\ 0 & 其他.\end{cases}$$

   a. 求常数 c 的值；
   b. 求 X_1 和 X_3 的边际联合概率密度函数；

c. 求 $\Pr\left(X_3 < \frac{1}{2} \mid X_1 = \frac{1}{4},\ X_2 = \frac{3}{4}\right)$.

2. 设三个随机变量 X_1，X_2 和 X_3 有混合型的联合分布，其概率函数/概率密度函数为

$$f(x_1,x_2,x_3) = \begin{cases} cx_1^{1+x_2+x_3}(1-x_1)^{(3-x_2-x_3)} & 0 < x_1 < 1 \text{ 且 } x_2,x_3 \in \{1,2\}, \\ 0 & \text{其他.} \end{cases}$$

（注意到 X_1 有连续型分布，X_2 和 X_3 有离散型分布.）（a）求常数 c 的值；（b）求 X_2 和 X_3 的边际联合概率函数；（c）求 $X_2 = 1$ 和 $X_3 = 1$ 的条件下 X_1 的条件概率密度函数.

3. 设三个随机变量 X_1，X_2 和 X_3 有连续型的联合分布，联合概率密度函数为：

$$f(x_1,x_2,x_3) = \begin{cases} ce^{-(x_1+2x_2+3x_3)} & \text{对于 } x_i > 1(i=1,2,3), \\ 0 & \text{其他.} \end{cases}$$

求（a）常数 c 的值；（b）X_1 和 X_3 的边际联合概率密度函数；（c）$\Pr(X_1 < 1 \mid X_2 = 2,\ X_3 = 1)$.

4. 设随机选取一点 (X_1,X_2,X_3)，即根据服从下面集合 S 上的均匀分布的概率密度函数：

$$S = \{(x_1,x_2,x_3) \mid 0 \leqslant x_i \leqslant 1, i = 1,2,3\}.$$

求a. $\Pr\left[(X_1 - \frac{1}{2})^2 + (X_2 - \frac{1}{2})^2 + (X_3 - \frac{1}{2}^2) \leqslant \frac{1}{4}\right]$

b. $\Pr\left(X_1^2 + X_2^2 + X_3^2 \leqslant 1\right)$.

5. 设一个电子系统有 n 个元件组成，每个元件的作用是相互独立的，且第 i 个元件正常工作的概率为 $p_i(i=1,\cdots,n)$. 一个串联系统正常工作的充分必要条件是每一个元件都正常工作. 一个并联系统正常工作的充分必要条件是至少有一个元件正常工作. 称系统正常工作的概率为系统的可靠性. 求这个系统的可靠性，（a）假设系统是串联的；（b）假设系统是并联的.

6. 设 n 个随机变量 $X_1,\cdots,X_n$ 组成概率函数为 f 的离散分布的随机样本. 求 $\Pr(X_1 = X_2 = \cdots = X_n)$ 的值.

7. 设 n 个随机变量 $X_1,\cdots,X_n$ 组成概率密度函数为 f 的连续分布的随机样本. 求这 n 个随机变量中至少有 k 个落在区间 $a \leqslant x \leqslant b$ 中的概率.

8. 设随机变量 X 的概率密度函数为：

$$f(x) = \begin{cases} \frac{1}{n!}x^n e^{-x} & \text{对于 } x > 0, \\ 0 & \text{其他.} \end{cases}$$

同时假设对于给定的 $X = x$（$x > 0$），n 个随机变量 $Y_1,\cdots,Y_n$ 是 i. i. d. 的，且它们中每一个的条件概率密度函数 g 为：

$$g(y \mid x) = \begin{cases} \frac{1}{x} & \text{对于 } 0 < y < x, \\ 0 & \text{其他.} \end{cases}$$

求（a）$Y_1,\cdots,Y_n$ 的边际联合概率密度函数；（b）对于任意给定的 $Y_1,\cdots,Y_n$ 的值，X 的条件概率密度函数.

9. 设 X 是有连续分布的随机变量，令 $X_1=X_2=X$.
   a. 证明 X_1 和 X_2 都具有连续型分布. b. 证明 $\boldsymbol{X}=(X_1,X_2)$ 不具有连续的联合分布.
10. 在例 3.7.8 中. 令 $\boldsymbol{X}=(X_1,\cdots,X_5)$，用一步直接计算在 $\boldsymbol{X}=\boldsymbol{x}$ 条件下 Z 的条件概率密度函数，可以看成同时观测到所有的 $\boldsymbol{X}$ 值.

157

11. 设 $X_1,\cdots,X_n$ 是互相独立的. 令 $k<n,i_1,\cdots,i_k$ 是 1 到 n 之间的不同的整数. 证明 $X_{i_1},\cdots,X_{i_k}$是互相独立的.
12. 设随机向量 X 被分成 3 个部分 $\boldsymbol{X}=(\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{W})$，假设 $\boldsymbol{X}$ 有连续的联合分布，其概率密度函数为 $f(\boldsymbol{y},\boldsymbol{z},\boldsymbol{w})$. 令 $g_1(\boldsymbol{y},\boldsymbol{z}\mid\boldsymbol{w})$ 为给定 $\boldsymbol{W}=\boldsymbol{w}$ 的条件下 $(\boldsymbol{Y},\boldsymbol{Z})$ 的条件概率密度函数，$g_2(\boldsymbol{y}\mid\boldsymbol{w})$ 为给定 $\boldsymbol{W}=\boldsymbol{w}$ 的条件下 $\boldsymbol{Y}$ 的条件概率密度函数. 证明：$g_2(\boldsymbol{y}\mid\boldsymbol{w})=\int g_1(\boldsymbol{y},\boldsymbol{z}\mid\boldsymbol{w})\mathrm{d}\boldsymbol{z}$.
13. 设对于任意的 z，在 $Z=z$ 的条件下，X_1，X_2，X_3 是条件独立的，且条件概率密度函数为 (3.7.11) 式中的 $g(x\mid z)$. 此外，Z 的边际概率密度函数为 (3.7.10) 式中的 f_0. 证明在给定 $(X_1,X_2)=(x_1,x_2)$ 的条件下的 X_3 的概率密度函数为 $\int_0^\infty g(x_3\mid z)g_0(z\mid x_1,x_2)\mathrm{d}z$，其中 g_0 如 (3.7.12) 式所定义.（即使不能得到积分的条件形式也可以证明此式）.
14. 考虑例 3.7.6 的情形. 假设观测到 $X_1=5$ 和 $X_2=7$.
   a. 计算在给定 $(X_1,X_2)=(5,7)$ 的条件下 X_3 的条件概率密度函数（可以利用习题 12 的结果.）
   b. 求在 $(X_1,X_2)=(5,7)$ 的条件下 $X_3>3$ 的条件概率. 并与例 3.7.3 得到的 $\Pr(X_3>3)$ 的值作比较. 思考：能否给出为什么条件概率比边际概率大的理由？

3.8 随机变量的函数

我们经常会发现在计算一个随机变量 X 的分布后，实际上要的是 X 的某些其他函数的分布. 例如，如果 X 是一个队伍中顾客接受服务的速率，则 $1/X$ 是平均等待时间，如果有了 X 的分布就能确定 $1/X$ 或者 X 的任意函数的分布. 如何来做就是本节的主题.

3.8.1 离散型随机变量

设一个随机变量 X 有离散型分布，其概率分布函数为 f，定义另一个随机变量 $Y=r(X)$ 为 X 的函数. 则可以如下从 f 得出 Y 的概率函数 g：对于 Y 的每一个可能值 y，

$$g(y)=\Pr(Y=y)=\Pr[r(X)=y]$$
$$=\sum_{x:r(x)=y}f(x).$$

3.8.2 连续型随机变量

如果一个随机变量 X 有连续型分布，则推导 X 的函数的概率分布的过程和离散型时是不一样的. 设 X 的概率密度函数为 f，另一个随机变量定义为 $Y=r(X)$. 对于任意的实数 y，Y 的分布函数 $G(y)$ 可以由下式给出：

$$G(y) = \Pr(Y \leqslant y) = \Pr\left[r(x) \leqslant y\right]$$

158

$$= \int_{\{x \mid r(x) \leqslant y\}} f(x)\,\mathrm{d}x.$$

如果随机变量 Y 也有连续型分布，可以由下式得出它的概率密度函数 g：

$$g(y) = \frac{\mathrm{d}G(Y)}{\mathrm{d}y}.$$

上式对于使 G 可微的任一点 y 成立.

例3.8.1　推导当 X 具有均匀分布时 X^2 的概率密度函数　设 X 具有在区间 $[-1,1]$ 上的均匀分布. 即

$$f(x) = \begin{cases} 1/2 & \text{对于 } -1 \leqslant x \leqslant 1, \\ 0 & \text{其他.} \end{cases}$$

确定随机变量 $Y = X^2$ 的概率密度函数.

因为 $Y = X^2$，所以 Y 必属于区间 $0 \leqslant Y \leqslant 1$. 因此，对于 Y 的满足 $0 \leqslant y \leqslant 1$ 任意值，Y 的分布函数 $G(y)$ 为：

$$G(y) = \Pr(Y \leqslant y) = \Pr(X^2 \leqslant y)$$

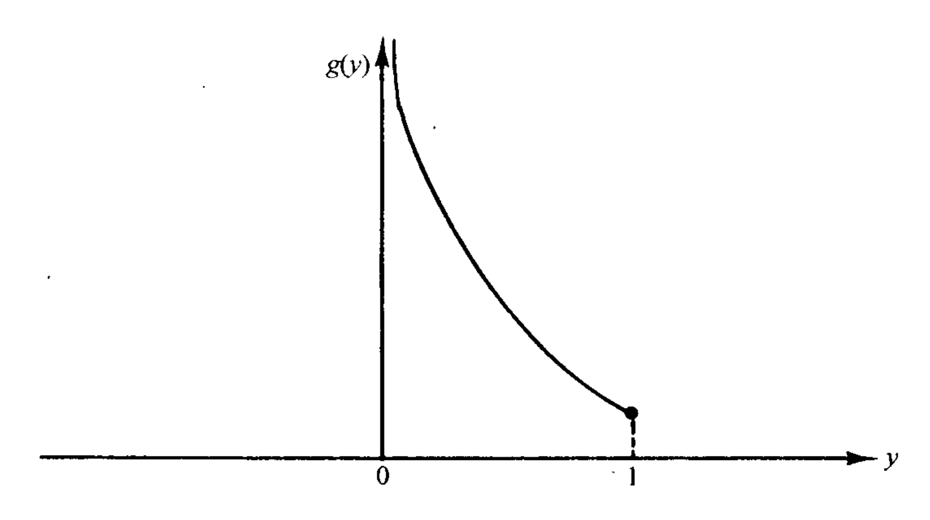
$$= \Pr(-y^{1/2} \leqslant X \leqslant y^{1/2})$$

$$\int_{-y^{1/2}}^{y^{1/2}} f(x)\,\mathrm{d}x = y^{1/2}.$$

对于 $0 < y < 1$，可以得到 Y 的概率密度函数 $g(y)$ 为：

$$g(y) = \frac{\mathrm{d}G(y)}{\mathrm{d}y} = \frac{1}{2y^{1/2}}.$$

Y 的这个概率密度函数如图3-20所示. 应当注意尽管 Y 只是有均匀分布的随机变量的平方，但是 Y 的概率密度函数在 $y = 0$ 的邻域内无界. ◀



159

图3-20　例3.8.1中 $Y = X^2$ 的概率密度函数

3.8.3 概率密度函数的直接推导

如果一个随机变量 X 具有连续分布，且 $Y=r(X)$，则 Y 不一定具有连续分布. 例如，设对于所有的 x 的值，$r(x)=c$，c 为常数，对于某个区间 $a\leqslant x\leqslant b$ 中 x 的所有值且 $\Pr(a\leqslant X\leqslant b)>0$，则 $\Pr(Y=c)>0$ 因为 $Y=c$ 的概率为正值. 所以这个分布是连续的. 为了计算这个例子中 Y 的分布，必须利用如上描述的方法来导出 Y 的概率. 然而，对于某些函数 r，Y 的分布是连续的，此时就不用先求概率函数而直接导出 Y 的概率密度函数.

设 $Y=r(X)$，且其中随机变量 X 必定落在区间(a,b)内，在该区间上 $r(x)$是严格单调递增的. 使 $\Pr(a<X<b)=1$ 的区间(a,b)可以是有界的、无界的或整个实轴. 当 x 在区间 $a<x<b$ 内变化时，$y=r(x)$在区间 $\alpha<y<\beta$ 中变化. 而且，对于满足 $\alpha<y<\beta$的任意值 y 都有唯一的 x，$a<x<b$，使得 $r(x)=y$. 如果记 x 的这个值为 $x=s(y)$，则函数 s 为 r 的逆函数. 即对于所有的满足 $a<x<b$ 和 $\alpha<y<\beta$ 的值 x 和 y，当且仅当 $x=s(y)$时 $y=r(x)$. 因为假设函数 r 在区间(a,b)是连续且严格单调递增的，所以逆函数 s 在区间(α,β)上也是连续且严格递增的. 因此，对于满足$\alpha<y<\beta$ 的任意值 y 有，

$$G(y)=\Pr(Y\leqslant y)=\Pr(r(X)\leqslant y)=\Pr[X\leqslant s(y)]=F[s(y)].$$

如果再假设 s 在区间(α,β)是可微函数，则 Y 的分布是连续的且对于 $\alpha<y<\beta$ 其概率密度函数$g(y)$为：

$$g(y)=\frac{\mathrm{d}G(y)}{\mathrm{d}y}=\frac{\mathrm{d}F[s(y)]}{\mathrm{d}y}=f[s(y)]\frac{\mathrm{d}s(y)}{\mathrm{d}y}. \tag{3.8.1}$$

换言之可由概率密度函数$f(x)$直接得出 Y 的概率密度函数 $g(y)$：只须用 y 的函数表示的 x 代替 x，且乘以 $\mathrm{d}x/\mathrm{d}y$.

类似地，如果 $Y=r(X)$且假设函数 r 在区间(a,b)上是连续且严格单调递减的，则当 X 在区间(a,b)变化时，Y 在区间(α,β)变化，且逆函数 s 在区间(α,β)上也是连续且严格单调递减的. 因此，对于 $\alpha<y<\beta$，

$$G(y)=\Pr[r(X)\leqslant y]=\Pr[X\geqslant s(y)]=1-F[s(y)].$$

如果再假设 s 在区间(α,β)是可微函数，则有

$$g(y)=\frac{\mathrm{d}G(y)}{\mathrm{d}y}=-f[s(y)]\frac{\mathrm{d}s(y)}{\mathrm{d}y}.$$

因为 s 是严格递减的，$\mathrm{d}s(y)/\mathrm{d}y<0$，所以 $g(y)$可以写成

$$g(y)=f[s(y)]\left|\frac{\mathrm{d}s(y)}{\mathrm{d}y}\right|. \tag{3.8.2}$$

160

因为当 r 和 s 严格递增时(3.8.1)式成立，r 和 s 严格递减时(3.8.2)式成立，所以可以将上面得到的结果概括为：

定理 3.8.1 设 X 是随机变量其概率密度函数为 f，且 $\Pr(a<X<b)=1$ 令 $Y=r(X)$，

且设 $r(x)$ 对于 $a<x<b$ 是连续的且严格单调递增或严格单调递减的. 同时设 $a<X<b$ 当且仅当 $\alpha<Y<\beta$, 令 $X=s(Y)$ 为对于 $\alpha<Y<\beta$ 的逆函数, 则 Y 的概率密度函数 g 为:

$$g(y)=\begin{cases} f[s(y)]\left|\dfrac{\mathrm{d}s(y)}{\mathrm{d}y}\right| & \text{对于 } \alpha<y<\beta, \\ 0 & \text{其他.} \end{cases} \tag{3.8.3}$$

■

例 3.8.2　细菌增长　大环境下细菌组织个数增长的典型例子是指数增长. 假设在 0 时刻把 v 个细胞放入一潭水中, 设 X 表示增长率. 经过时间 t 后, 我们期望细胞的数目为 ve^{Xt}. 假设 X 是未知的但具有连续分布, 其概率密度函数为:

$$f(x)=\begin{cases} 3(1-x)^2 & \text{对于 } 0<x<1, \\ 0 & \text{其他.} \end{cases}$$

我们感兴趣的是对于已知 v 和 t 值时 $Y=ve^{Xt}$ 的分布. 具体地, 令 $v=10$ 和 $t=5$, 因此 $r(x)=10e^{5x}$.

在这个例子中 $\Pr(0<X<1)=1$, 且 x 的函数 r 对于 $0<x<1$ 是连续严格递增的. 当 x 在区间 $(0,1)$ 变化时, $y=r(x)$ 在区间 $(10,10e^5)$ 变化. 而且对于 $10<y<10e^5$, 逆函数是 $s(y)=\log(y/10)/5$. 因此, 对于 $10<y<10e^5$, 有

$$\frac{\mathrm{d}s(y)}{\mathrm{d}y}=\frac{1}{5y}.$$

由 (3.8.3) 式知 $g(y)$ 为:

$$g(y)=\begin{cases} \dfrac{3(1-\ln(y/10)/5)^2}{5y} & \text{对于 } 10<y<10e^5, \\ 0 & \text{其他.} \end{cases}$$

◀

3.8.4　概率积分变换

假设随机变量 X 具有连续分布函数 F, 令 $Y=F(X)$. 称这个从 X 到 Y 的变换为**概率积分变换**. 我们将证明 Y 的分布一定是区间 $[0,1]$ 上的均匀分布.

首先, 因为 F 是随机变量的分布函数, 所以 $0\leqslant F(x)\leqslant 1$, $-\infty<X<\infty$. 因此, $\Pr(Y<0)=\Pr(Y>1)=0$. 因为 F 是连续的, 所以使得对于每个 $y\in(0,1)$, $F(x)=y$ 的 x 的集合是非空的闭有界区间 $[x_0,x_1]$. 令 $F^{-1}(y)$ 表示这个区间的左端点 x_0, 在第 3.2 节中称它为 F 的 y 分位数. 因此 $Y\leqslant y$ 当且仅当 $X\leqslant x_1$. 令 G 表示 Y 的分布函数. 则

161

$$G(y)=\Pr(Y\leqslant y)=\Pr(X\leqslant x_1)=F(x_1)=y.$$

因此, 对于 $0<y<1$, $G(y)=y$. 因为这个函数是区间 $[0,1]$ 上的均匀分布的分布函

数，这个均匀分布就是 Y 的分布.

现在假设 X 是具有连续分布函数 F 的随机变量，且 G 是实轴上的连续分布函数. 我们想构造随机变量 $Z=r(X)$ 使其分布函数为 G.

因为 G 是连续的，所以对于区间 $0<y<1$ 上 y 的每个值，使得 $G(z)=y$ 的 z 的集合是有界闭区间 $[z_0,z_1]$. 令 $G^{-1}(y)$ 表示这个区间的左端点 z_0，z_0 为分布 G 的 y 分位数. 现在，定义以下的随机变量 Z:

$$Z=G^{-1}[F(X)].$$

这样，当且仅当 $F(X)\leqslant G(z)$ 时 $Z\leqslant z$. 为了证明 Z 的分布函数正好是 G，我们注意到对于使得 $0<G(z)<1$ 的每个数 z 有，

$$\Pr(Z\leqslant z)=\Pr[F(X)\leqslant G(X)].$$

由概率积分变换知 $F(X)$ 有均匀分布，因此

$$\Pr[F(X)\leqslant G(X)]=G(z).$$

从而，$\Pr(Z\leqslant z)=G(z)$，即意味着 Z 的分布函数如所要求的正是 G.

3.8.5 模拟

1. 伪随机数

大多数做统计分析的计算机软件包都能生成我们所称的*伪随机数*. 这些数即使是由确定的算法生成的，它们仍具有一个随机样本所拥有的一些性质. 这些程序的最基本的是能生成具有区间[0,1]上均匀分布的伪随机数. 我们把这些函数称为*均匀的伪随机数发生器*. 均匀的伪随机数发生器必须具备如下所述的重要特征：它生成的数相当均匀地分布在区间[0,1]上，且必须看起来是独立随机变量的观测值. 最后一项特征比较难于用语言精确地描述. 一个例子是一个序列看起来不是独立随机变量的观测值，而它是完全的平均的间隔开来的. 另一个例子是有下述状态的序列，假设我们看序列 $X_1,X_2,\cdots$一次看一个，并且每次发现一个 $X_i>0.5$，我们就记录下下一个数 X_{i+1}. 如果我们记录的数的序列不是近似地均匀分布在[0,1]上的，则原始的序列就不像是区间[0,1]上均匀分布的独立随机变量的观测值. 原因是给定 $X_i>0.5$，X_{i+1} 的条件分布应该是区间[0, 1]上的均匀分布. 162

2. 生成有既定分布的伪随机数

可以用一个均匀的伪随机数发生器来生成一个具有任意给定的连续分布函数 G 的随机变量 Y 的值. 如果随机变量 X 是区间[0,1]上的均匀分布，且分位数函数 G^{-1} 的定义如前，则由概率积分变换可知随机变量 $Y=G^{-1}(X)$ 的分布函数就是 G. 因此如果 X 的值是由均匀的伪随机数发生器生成的，则相应的 Y 值具备我们需要的性质. 如果 n 个独立的值 $X_1,\cdots,X_n$ 是由发生器生成的，则相应的值 $Y_1,\cdots,Y_n$ 将形成一个由分布函数 G 得到的容量为 n 的随机样本.

例 3.8.3 从给定的概率密度函数生成独立的值 假设用一个均匀的伪随机数发生器生成概率密度函数是 g 的 3 个独立值，g 的定义如下：

$$g(y)=\begin{cases}\frac{1}{2}(2-y) & \text{当 } 0<y<2,\\ 0 & \text{其他}.\end{cases}$$

对于 $0<y<2$，给定分布的分布函数 G 是

$$G(y)=y-\frac{y^2}{4}.$$

并且，当 $0<x<1$，可以通过解方程 $x=G(y)$ 求得反函数 $y=G^{-1}(x)$. 结果是

$$y=G^{-1}(x)=2[1-(1-x)^{1/2}]. \tag{3.8.4}$$

下一步就是用发生器生成 3 个均匀的伪随机数 x_1，x_2 和 x_3. 设 3 个生成的值是

$$x_1=0.4125, x_2=0.0894, x_3=0.8302.$$

当将这些值 x_1，x_2 和 x_3 依次代入(3.8.4)式我们就得到 y 的值 $y_1=0.47$，$y_2=0.09$ 和 $y_3=1.18$. 这些值就被看作是概率密度函数为 g 的分布的 3 个独立随机变量的观测值. ◀

如果 G 是一个一般的分布函数，则有一个类似于概率积分变换的方法可以把一个均匀分布的随机变量变成分布函数为 G 的随机变量. 参看本节习题 12. 还有一些其他的比概率积分变换更快更精确的计算方法可由指定的分布生成值. 这些内容在 Kennedy 和 Gentle(1980)和 Rubinstein(1981)的书里都有讨论. 第 11 章包含了怎样用模拟来解决统计问题的技巧和例题. 163

3.8.6 小结

我们学习了几种求一个随机变量的函数的分布的方法. 对于一个概率密度函数为 f 的连续分布的随机变量 X，如果 r 是严格单调增或严格单调降的且有可微逆函数[也就是 $s(r(x))=x$，且 s 是可微的]，则 $Y=r(x)$ 的概率密度函数是 $g(y)=f(s(y))|ds(y)/dy|$. 一个特别的变换，称为概率积分变换，允许我们通过 $Y=G^{-1}(X)$ 把一个具有区间[0, 1]上均匀分布的随机变量 X 变换成具有任意的连续分布函数 G 的随机变量 Y. 这种方法可以和均匀的伪随机数发生器相结合生成具有任意连续分布的随机变量.

习题

1. 设随机变量 X 的概率密度函数是

$$f(x)=\begin{cases}3x^2 & \text{当 } 0<x<1,\\ 0 & \text{其他}.\end{cases}$$

且设 $Y=1-X^2$. 求 Y 的概率密度函数.

2. 设随机变量 X 以等概率取 $-3,-2,-1,0,1,2,3$ 这七个值中的各个值. 求 $Y=X^2-X$ 的概率函数.

3. 设随机变量 X 的概率密度函数是

$$f(x)=\begin{cases}\frac{1}{2}x & \text{当 } 0<x<2,\\ 0 & \text{其他.}\end{cases}$$

且 $Y=X(2-X)$. 求 Y 的概率密度函数和分布函数.

4. 设 X 的概率密度函数是习题 3 中所给出的函数，求 $Y=4-X^3$ 的概率密度函数.

5. 设随机变量 X 的概率密度函数是 f，且 $Y=aX+b(a\neq 0)$. 证明 Y 的概率密度函数是

$$G(Y)=\frac{1}{|a|}f\left(\frac{y-b}{a}\right)\quad \text{对于 } -\infty<y<\infty.$$

6. 设 X 的概率密度函数是习题 3 中所给出的函数，求 $Y=3X+2$ 的概率密度函数.

7. 设随机变量 X 具有区间 $[0,1]$ 上的均匀分布，求下列变量的概率密度函数 (a) X^2, (b) $-x^3$, (c) $X^{1/2}$.

8. 设随机变量 X 的概率密度函数如下：

$$f(x)=\begin{cases}\mathrm{e}^{-x} & \text{当 } x>0,\\ 0 & \text{当 } x\leqslant 0.\end{cases}$$

求 $Y=X^{1/2}$ 的概率密度函数.

9. 设随机变量 X 具有区间 $[0,1]$ 上的均匀分布，试构造随机变量 $Y=r(X)$ 使得它的概率密度函数是

$$g(y)=\begin{cases}\frac{3}{8}y^2 & \text{当 } 0<y<2,\\ 0 & \text{其他.}\end{cases}$$

10. 设 X 的概率密度函数是习题 3 中所给出的函数，试构造随机变量 $Y=r(X)$ 使得它的概率密度函数是由习题 9 所给出的函数 g.

11. 解释怎样从均匀的伪随机数发生器来生成其分布的概率密度函数是

$$g(y)=\begin{cases}\frac{1}{2}(2y+1) & \text{当 } 0<y<1,\\ 0 & \text{其他.}\end{cases}$$

的四个独立的值.

12. 设 F 是任意一个分布函数（不一定是离散，不一定是连续，也不一定是其中任意一个）. 设 F^{-1} 是 3.3 节所定义的分位数函数，X 具有在区间 $[0,1]$ 上的均匀分布. 定义 $Y=F^{-1}(X)$. 证明：Y 的分布函数是 F.

提示：在两种情况下计算 $\Pr(Y\leqslant y)$. 第一，y 是唯一的满足 $F(x)=F(y)$ 的 x 的值. 第二，在 x 的取值的全区间上 $F(x)=F(y)$.

13. 设 Z 是对排队中的消费者提供服务的速度. 并设 Z 具有概率密度函数

$$f(z)=\begin{cases}2e^{-2z} & 当 z>0,\\ 0 & 当 z\leqslant 0.\end{cases}$$

求平均等待时间 $T=1/Z$ 的概率密度函数.

14. 设随机变量 X 具有区间 $[a,b]$ 上的均匀分布. 证明：$cX+d$ 具有区间 $[ca+d,cb+d]$ 上的均匀分布.

164

3.9 两个或两个以上随机变量的函数

当我们观察的数据是由多个随机变量的值组成时，我们需要综合这些观测值从而能够聚焦在这些数据的信息上. 综合包括构造一个或几个关于这些随机变量的能够捕获大量信息的函数. 在本节中，我们将叙述用来求两个或多个随机变量的函数的分布的技巧.

3.9.1 具有离散型联合分布的变量

设 n 个具有离散型联合分布的随机变量 $X_1,\cdots,X_n$ 的联合概率函数为 f，且定义以下 m 个关于这 n 个随机变量的函数 $Y_1,\cdots,Y_m$：

$$\begin{aligned}Y_1&=r_1(X_1,\cdots,X_n),\\ Y_2&=r_2(X_1,\cdots,X_n),\\ &\ \ \vdots\qquad\quad \vdots\\ Y_m&=r_m(X_1,\cdots,X_n).\end{aligned}$$

对于 m 个随机变量 $Y_1,\cdots,Y_m$ 的给定值 $y_1,\cdots,y_m$，令 A 表示所有满足

$$\begin{aligned}r_1(x_1,\cdots,x_n)&=y_1,\\ r_2(x_1,\cdots,x_n)&=y_2,\\ &\ \ \vdots\qquad\quad \vdots\\ r_m(x_1,\cdots,x_n)&=y_m.\end{aligned}$$

的点的集合. 则 $Y_1,\cdots,Y_m$ 的联合概率函数 g 在给定点 $y_1,\cdots,y_m$ 的值由下式给出

$$g(y_1,\cdots,y_m)=\sum_{(x_1,\cdots,x_n)\in A} f(x_1,\cdots,x_n).$$

3.9.2 具有连续型联合分布的变量

如果 $X_1,\cdots,X_n$ 的联合分布是连续的，则必须使用有别于离散型随机变量的一些方法来获得 $X_1,\cdots,X_n$ 的函数的分布或者两个或多个这样函数的联合分布. 如果 $X_1,\cdots,X_n$ 的联合概率密度函数是 $f(x_1,\cdots,x_n)$ 且如果 $Y=r(X_1,\cdots,X_n)$，则可由基本的原则求出 Y 的分布函数 $G(y)$. 对于每个给定的值 $y(-\infty<y<\infty)$，令 A_y 是 $\mathbf{R}^n$ 的包含所有满足 $r(y_1,\cdots,y_m)\leqslant y$ 的点 $(x_1,\cdots,x_n)$ 的集合. 则

$$G(y)=\Pr(Y\leqslant y)=\Pr[r(X_1,\cdots,X_n)\leqslant y]=\int\underset{A_y}{\cdots}\int f(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

如果 Y 的分布也是连续的，则 Y 的概率密度函数可以由对分布函数 $G(y)$ 求微分得到. 165

随机样本的最大值最小值的分布

关于一个随机样本的普通的概述由最大最小观测值组成，该函数告诉我们这些数据是如何散布的. 设变量 $X_1,\cdots,X_n$ 构成取自概率密度函数是 f 且分布函数是 F 的分布且容量为 n 的随机样本. 该随机样本的最大值 Y_n 定义为

$$Y_n=\max\{X_1,\cdots,X_n\}.$$

我们来确定 Y_n 的分布函数 G_n 和概率密度函数 g_n.

对于 y 的任意一个给定的值（$-\infty<y<\infty$），

$$\begin{aligned}G_n(y)&=\Pr(Y_n\leqslant y)=\Pr(X_1\leqslant y,X_2\leqslant y,\cdots,X_n\leqslant y)\\&=\Pr(X_1\leqslant y)\Pr(X_2\leqslant y)\cdots\Pr(X_n\leqslant y)\\&=F(y)F(y)\cdots F(y)=[F(y)]^n.\end{aligned}$$

其中第三个等式是由 X_i 相互独立得到的，第四个等式是由所有的 X_i 有相同的概率分布 F 得到的. 从而 $G_n(y)=[F(y)]^n$.

可以由对分布函数 G_n 求微分得到 Y_n 的概率密度函数 g_n. 结果是

$$g_n(y)=n[F(y)]^{n-1}f(y)\qquad 当 -\infty<y<\infty.$$

166

随机样本中最小值 Y_1 定义为：

$$Y_1=\min\{X_1,\ \cdots,\ X_n\}.$$

对于 y 的任意一个给定的值($-\infty<y<\infty$)，可由下式得到 Y_1 的分布函数 G_1：

$$\begin{aligned}G_1(y)&=\Pr(Y_1\leqslant y)=1-\Pr(Y_1>y)\\&=1-\Pr(X_1>y,X_2>y,\cdots,X_n>y)\\&=1-\Pr(X_1>y)\Pr(X_2>y)\cdots\Pr(X_n>y)\\&=1-[1-F(y)][1-F(y)]\cdots[1-F(y)]\\&=1-[1-F(y)]^n.\end{aligned}$$

从而 $G_1(y)=1-[1-F(y)]^n$.

可以由对分布函数 G_1 求微分得到 Y_1 的概率密度函数 g_1. 结果是

$$g_1(y)=n[1-F(y)]^{n-1}f(y)\qquad 对于 -\infty<y<\infty 成立$$

最后，我们来演示怎样由 Y_1 和 Y_n 的联合分布来求 $X_1,\cdots,X_n$ 的两个和多个函数的联合分布的方法. 对于每对满足 $-\infty<y_1<y_n<\infty$ 的值 (y_1,y_n)，事件 $Y_1\leqslant y_1$ 且 $Y_n\leqslant y_n$ 和事件 $Y_n\leqslant y_n$ 与 $Y_1>y_1$ 的余集之交相同. 则

$$\begin{aligned}G(y_1,y_n)&=\Pr(Y_1\leqslant y_1\ 且\ Y_n\leqslant y_n)\\&=\Pr(Y_n\leqslant y_n)-\Pr(Y_n\leqslant y_n\ 且\ Y_1>y_1)\\&=\Pr(Y_n\leqslant y_n)-\Pr(y_1<X_1\leqslant y_n,y_1<X_2\leqslant y_n,\cdots,y_1<X_n\leqslant y_n)\\&=G_n(y_n)-\prod_{i=1}^{n}\Pr(y_1<X_i\leqslant y_n)\\&=[F(y_n)]^n-[F(y_n)-F(y_1)]^n.\end{aligned}$$

可由下列关系式得到 Y_1 和 Y_n 的二元联合概率密度函数 g

$$g(y_1,y_n)=\frac{\partial^2 G(y_1,y_n)}{\partial y_1 \partial y_n}.$$

这样，当 $-\infty<y_1<y_n<\infty$ 时，

$$g(y_1,y_n)=n(n-1)[F(y_n)-F(y_1)]^{n-2}f(y_1)f(y_n). \tag{3.9.1}$$

并且对于 y_1 和 y_n 的其他的值，有 $g(y_1,y_n)=0$. 在本节的最后我们将返回到联合概率密度函数.

3.9.3 多元概率密度函数的变换

我们继续设随机变量 $X_1,\cdots,X_n$ 有连续型联合分布，它们的联合概率密度函数是 f. 我们来考虑 n 个新的随机变量 $Y_1,\cdots,Y_n$ 的联合概率密度函数 g，其中 $Y_1,\cdots,Y_n$ 的定义如下：

$$\begin{aligned} Y_1 &= r(X_1,\cdots,X_n), \\ Y_2 &= r_2(X_1,\cdots,X_n), \\ &\vdots \quad\quad \vdots \\ Y_n &= r_n(X_1,\cdots,X_n). \end{aligned} \tag{3.9.2}$$

我们要对函数 $r_1,\cdots,r_n$ 做一定的假设. 首先设 S 是 $\mathbf{R}^n$ 的满足 $\Pr[(X_1,\cdots,X_n)\in S]=1$ 的子集. 子集 S 可以是整个空间 $\mathbf{R}^n$，然而如果 $X_1,\cdots,X_n$ 的概率分布集中在 $\mathbf{R}^n$ 的某个真子集上，则可选这个更小的集合作 S. 令 T 表示 $\mathbf{R}^n$ 的子集，它是由变换(3.9.2)的 n 个方程决定的变换下 S 的像集. 换言之，当$(X_1,\cdots,X_n)$的值在集合 S 中变化时，$(Y_1,\cdots,Y_n)$的值在集合 T 中变化. 并且进一步假定从 S 到 T 的变换是一对一的变换，换言之，对于集合 T 中$(Y_1,\cdots,Y_n)$的每一个值存在集合 S 中唯一的一个$(X_1,\cdots,X_n)$的满足(3.9.2)的 n 个方程的值与之对应.

167 从最后一个假定可以得到 S 中的点$(x_1,\cdots,x_n)$和 T 中的点$(y_1,\cdots,y_m)$存在一一对应关系. 因此，对于 T 中的$(y_1,\cdots,y_m)$我们可以把(3.9.2)中的方程转化得到具有下列形式的新的方程

$$\begin{aligned} x_1 &= s_1(y_1,\cdots,y_n), \\ x_2 &= s_2(y_1,\cdots,y_n), \\ &\vdots \quad\quad \vdots \\ x_n &= s_n(y_1,\cdots,y_n). \end{aligned} \tag{3.9.3}$$

其次，假设对于 $i=1,\cdots,n$ 且 $j=1,\cdots,n$，在每一点$(y_1,\cdots,y_m)\in T$ 偏微分$\partial s_i/\partial y_j$ 都存在. 在该种假设下就可以构造下列矩阵 $\boldsymbol{J}$：

$$\boldsymbol{J}=\det\begin{pmatrix} \dfrac{\partial s_1}{\partial y_1} & \cdots & \dfrac{\partial s_1}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial s_n}{\partial y_1} & \cdots & \dfrac{\partial s_n}{\partial y_n} \end{pmatrix}.$$

称这个矩阵为方程(3.9.3)所确定的变换的雅可比矩阵.

可以通过高等微积分中多元函数变换积分变量的方法得到这 n 个随机变量 $Y_1,\cdots,Y_n$的联合概率密度 g. 此处就不列出具体的推导，只列出结果如下：

$$g(y_1,\cdots,y_m)=\begin{cases}f(s_1,\cdots,s_n)\,|\boldsymbol{J}| & \text{当}(y_1,\cdots,y_n)\in T,\\ 0 & \text{其他}.\end{cases} \tag{3.9.4}$$

在方程(3.9.4)中，$|\boldsymbol{J}|$表示矩阵 $\boldsymbol{J}$ 的绝对值. 这样，联合概率密度 $g(y_1,\cdots,y_n)$就可以通过把联合概率密度函数 $f(x_1,\cdots,x_n)$ 中的每个 x_i 用关于 $y_1,\cdots,y_n$ 的表达式 $s_i(y_1,\cdots,y_n)$代替，再乘以$|\boldsymbol{J}|$的值得到.

注：雅可比变换是反函数求导的推广. 方程(3.8.2)和(3.9.4)是非常相似的. 前者给出了单个随机变量的单个函数的概率密度函数. 事实上，如果在(3.9.4)中取 $n=1$，则 $J=\mathrm{d}s_1(y_1)/\mathrm{d}y_1$，则方程(3.9.4)就变得和方程(3.8.2)一样. 雅可比变换仅仅是从一个变量的单个函数的反函数求导推广到 n 个变量的 n 元函数.

例 3.9.1 两个随机变量的积和商的联合概率密度函数 设两个随机变量 X_1 和 X_2 有一个连续的联合分布，其联合概率密度函数如下：

$$f(x_1,x_2)=\begin{cases}4x_1x_2 & \text{当}\,0<x_1<1\,\text{且}\,0<x_2<1,\\ 0 & \text{其他}.\end{cases}$$

下面来求两个新随机变量 Y_1 和 Y_2 的联合概率密度，它们的定义如下：

$$Y_1=\frac{X_1}{X_2}\text{和}\ Y_2=X_1X_2. \tag{3.9.5}$$

168

在方程(3.9.5)中我们解出 X_1 和 X_2 作为 Y_1 和 Y_2 的函数的表达式是

$$X_1=(Y_1Y_2)^{1/2}\text{和}\ X_2=\left(\frac{Y_2}{Y_1}\right)^{1/2}. \tag{3.9.6}$$

如果 S 表示满足 $0<x_1<1$ 且 $0<x_2<1$ 的点(x_1,x_2)的集合，则 $\Pr[(x_1,x_2)\in S]=1$. 而且，从关系式(3.9.5)和(3.9.6)中可以看出条件 $0<X_1<1$ 且 $0<X_2<1$ 和条件 $Y_1>0$，$Y_2>0$，$Y_1Y_2<1$ 且$(Y_2/Y_1)<1$ 是等价的. 因此当 X_1 和 X_2 的值(x_1,x_2)在 S 中变化时，Y_1 和 Y_2 的值(y_1,y_2)将在包含 $y_1>0$，$y_2>0$，$(y_1y_2)^{1/2}<1$ 且$(y_2/y_1)^{1/2}<1$ 的所有点的集合 T 中变化. 集合 S 和 T 的图像如图 3-21 所示.

由(3.9.5)中的方程所定义的变换，或者，等价的说，由(3.9.6)中的方程规定的集合 S 和 T 中的点的一一对应关系. 对于$(y_1,y_2)\in T$，变换由下面的关系给出：

$$x_1=s_1(y_1,y_2)=(y_1y_2)^{1/2},$$

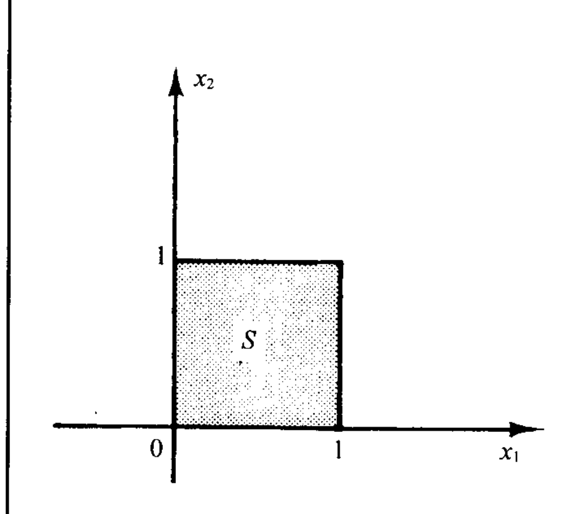
$$x_2=s_2(y_1,y_2)=\left(\frac{y_2}{y_1}\right)^{1/2}.$$

对于这些关系式，

$$\frac{\partial s_1}{\partial y_1}=\frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2},\frac{\partial s_1}{\partial y_2}=\frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2},$$

169

$$\frac{\partial s_2}{\partial y_1}=-\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2},\ \frac{\partial s_2}{\partial y_2}=\frac{1}{2}\left(\frac{1}{y_1y_2}\right)^{1/2}.$$



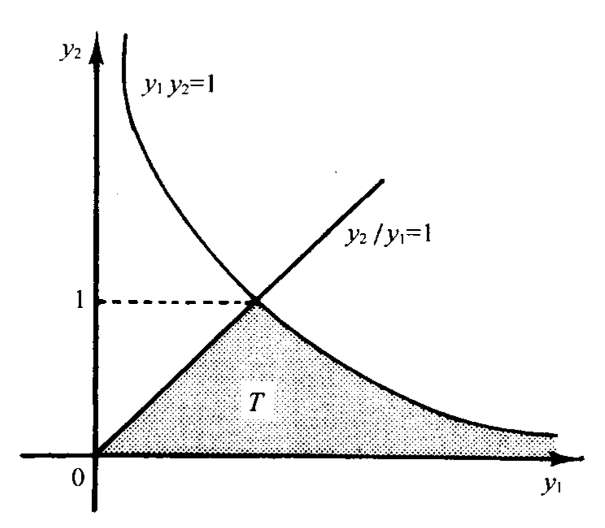


图 3-21　例 3.9.1 中的集合 S 和 T

因此，

$$J=\det\begin{pmatrix}\frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2}\\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2} & \frac{1}{2}\left(\frac{1}{y_1y_2}\right)^{1/2}\end{pmatrix}=\frac{1}{2y_1}.$$

由于在整个集合 T 上 $y_1>0$，所以 $|J|=1/(2y_1)$.

可以直接从方程(3.9.4)以下列方式得到联合概率密度函数 $g(y_1,y_2)$：在表达式 $f(x_1,x_2)$ 中，用 $(y_1y_2)^{1/2}$ 代替 x_1，用 $(y_2/y_1)^{1/2}$ 代替 x_2，再乘以 $|J|=1/(2y_1)$ 即得. 因此，

$$g(y_1,y_2)=\begin{cases}2\left(\frac{y_2}{y_1}\right) & \text{当}(y_1,y_2)\in T,\\ 0 & \text{其他}.\end{cases}$$

◀

例 3.9.2　排队中的服务时间　设服务系统最多只能同时为一个顾客服务，令 X 表示该服务系统为某一指定顾客服务的时间，令Y表示服务系统运行的速率. 给定

$Y=y$关于 X 的条件分布的一个普遍的模型，即对给定 $Y=y$，X 的条件概率密度是

$$g_1(x|y)=\begin{cases}ye^{-xy} & 当\ x>0,\\ 0 & 其他.\end{cases}$$

设 Y 的概率密度函数是 $f_2(y)$. 则 (X,Y) 的联合概率密度函数是 $g_1(x|y)f_2(y)$. 一个相对于平均而言度量顾客接受服务有多快的测度是 $Z=XY$. 例如，$Z=1$ 对应于平均的服务时间，而 $Z>1$ 意味着这位顾客接受的服务时间比平均时间长，$Z<1$ 意味着这位顾客接受服务的时间要比平均时间短. 如果要求 Z 的分布，我们就直接用刚才演示过的方法计算 (Z,Y) 的联合概率密度，将联合概率密度函数对 y 积分得到 Z 的边际概率密度函数. 然而由于条件 $Y=y$ 允许我们将 Y 看作常数 y，将给定 $Y=y$ 关于 X 的条件分布变换成给定 $Y=y$ 关于 Z 的条件分布会更加简单. 因为 $X=Z/Y$，逆变换是 $x=s(z)$ 其中 $s(z)=z/y$. 它的导数是 $1/y$，给定 $Y=y$，Z 的条件概率密度是

$$h_1(z/y)=\frac{1}{y}g_1\left(\frac{z}{y}\,\Big|\,y\right).$$

因为 Y 是速率，$Y\geqslant 0$ 且 $X=Z/Y>0$ 当且仅当 $Z>0$. 因此

$$h_1(z/y)=\begin{cases}e^{-z} & 当\ z>0,\\ 0 & 其他.\end{cases}\tag{3.9.7}$$

注意到 h_1 是不依赖于 y 的，因此 Z 和 Y 是独立的，h_1 是 Z 的边际概率密度函数. 读者可以在习题 17 中验证这些结论. ◀

注：消除依赖性. 在例 3.9.2 中公式 $Z=XY$ 使得看上去 Z 是依赖于 Y 的. 然而，事实上用 Y 乘以 X 就消除了 X 已有的对 Y 的依赖性，使得结果独立于 Y. 这种消除一个变量对于另一个变量的依赖性的变换的类型是一个非常有用的求随机变量的变换的边际分布的技巧.

170

3.9.4 线性变换

为了进一步解释我们得到的结果，继续假设 n 个随机变量 $X_1,\cdots,X_n$ 有连续型联合分布，它的联合概率密度函数为 f，现在来考虑由随机变量 $X_1,\cdots,X_n$ 的线性变换得到的新随机变量 $Y_1,\cdots,Y_n$ 的问题.

设 $\boldsymbol{A}$ 是一个给定的 $n\times n$ 阶矩阵，并且设 n 个随机变量 $Y_1,\cdots,Y_n$ 是通过下式定义的：

$$\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix}=\boldsymbol{A}\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}.\tag{3.9.8}$$

这样，每个随机变量 Y_i 都是随机变量 $X_1,\cdots,X_n$ 的一个线性组合. 我们同样还假设矩

阵 $\boldsymbol{A}$ 是非奇异的，从而矩阵 $\boldsymbol{A}^{-1}$ 存在. 由此，方程(3.9.8)中的变换是整个空间 $\mathbf{R}^n$ 到它自身上的一一变换. 在每一点 $(y_1,\cdots,y_n)\in\mathbf{R}^n$，通过下面方程可以将逆变换表示出来：

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{A}^{-1}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \tag{3.9.9}$$

将方程(3.9.9)所定义的变换的雅可比矩阵简记为 $\boldsymbol{J}=\det \boldsymbol{A}^{-1}$. 并且从行列式理论中知道

$$\det \boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}}.$$

因此，在每一点 $(y_1,\cdots,y_n)\in\mathbf{R}^n$，可以用下面方法得到 $y_1,\cdots,y_n$ 的联合概率密度 $g(y_1,\cdots,y_n)$：首先，对于 $i=1,\cdots,n$，$f(x_1,\cdots,x_n)$ 中的每个分量 x_i 用向量

$$\boldsymbol{A}^{-1}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

的第 i 个分量来代替，再除以 $|\det \boldsymbol{A}|$ 即得结果.

用向量表示，令

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{和}\ \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

并且我们分别用 $f(\boldsymbol{x})$ 和 $g(\boldsymbol{y})$ 来表示 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_n$ 的联合概率密度，则

171

$$g(\boldsymbol{y}) = \frac{1}{|\det \boldsymbol{A}|} f(\boldsymbol{A}^{-1}\boldsymbol{y}) \qquad \text{当}\ \boldsymbol{y}\in\mathbf{R}^n. \tag{3.9.10}$$

3.9.5　两个随机变量的和

设两个随机变量 X_1 和 X_2 有一个给定的联合概率密度函数 f，我们来求随机变量 $Y=X_1+X_2$ 的联合概率密度函数.

为方便起见，令 $Z=X_2$. 则从 X_1 和 X_2 到 Y 和 Z 的变换是一一的线性变换. 由下述方程表出逆变换：

$$X_1 = Y - Z,$$
$$X_2 = Z.$$

该变换的系数矩阵 A^{-1} 是

$$A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

从而

$$\det A^{-1} = \frac{1}{\det A} = 1.$$

从方程(3.9.10)就可以得到 Y 和 Z 在给定的点 (y,z) 的联合概率密度 g_0 为

$$g_0(y,z) = f(y-z,z).$$

因此，就可以通过下式得到 Y 的边际概率密度 g:

$$g(y) = \int_{-\infty}^{\infty} f(y-z,z)\,dz \qquad 当 -\infty < y < \infty. \tag{3.9.11}$$

如果开始时用 $Z = X_1$ 定义 Z，我们就得到概率密度函数 g 的另一个等价形式:

$$g(y) = \int_{-\infty}^{\infty} f(z,y-z)\,dz \qquad 当 -\infty < y < \infty. \tag{3.9.12}$$

如果 X_1 和 X_2 是独立的随机变量，它们的边际概率密度函数是 f_1 和 f_2，则对于 x_1 和 x_2 的所有的值有，$f(x_1,x_2) = f_1(x_1)f_2(x_2)$.

由方程(3.9.11)和(3.9.12)可知 $Y = X_1 + X_2$ 的概率密度函数由下列任意一个关系式给出

$$g(y) = \int_{-\infty}^{\infty} f_1(y-z)f_2(z)\,dz \qquad 当 -\infty < y < \infty, \tag{3.9.13}$$

$$或 \quad g(y) = \int_{-\infty}^{\infty} f_1(z)f_2(y-z)\,dz \qquad 当 -\infty < y < \infty. \tag{3.9.14}$$

将由方程(3.9.13)或(3.9.14)所确定的概率密度函数 g 称为概率密度函数 f_1 和 f_2 的卷积.

172

例 3.9.3 求卷积的概率密度函数 设 X_1 和 X_2 是独立同分布的随机变量，它们的概率密度函数如下:

$$f(x) = \begin{cases} e^{-x} & 当 x \geq 0, \\ 0 & 其他. \end{cases}$$

我们来求随机变量 $Y = X_1 + X_2$ 的概率密度函数 g.

由方程(3.9.13)或(3.9.14)有,

$$g(y) = \int_{-\infty}^{\infty} f(y-z)f(z)\,dz \qquad 当 -\infty < y < \infty. \tag{3.9.15}$$

由于，当 $x<0$ 时 $f(x) = 0$，则当 $z<0$ 或 $z>y$ 方程(3.9.15)中的被积函数为零. 因此当 $y \geq 0$ 时,

$$g(y) = \int_0^y f(y-z)f(z)\,dz = \int_0^y e^{-(y-z)}e^{-z}dz$$

$$= \int_0^y e^{-y}dz = ye^{-y}.$$

当 $y<0$ 时,

$$g(y) = 0.$$

◀

例 3.9.4　投资组合　假设一个投资者既想买股票又想买债券. 令 X_1 表示一年后股票的价值，X_2 表示一年后债券的价值，并设 X_1 和 X_2 相互独立. 设 X_1 在区间 [1000,4000] 上有均匀分布，X_2 在区间 [800,1200] 上有均匀分布. 和 $Y = X_1 + X_2$ 是由股票和债券组成的组合在一年后的价值. 我们来求 Y 的概率密度函数. 方程 (3.9.14) 的函数 $f_1(z)f_2(y-z)$ 是

$$f_1(z)f_2(y-z) = \begin{cases} 8.333 \times 10^{-7} & \text{当 } 1000 \leqslant z \leqslant 4000 \text{ 且 } 800 \leqslant y-z \leqslant 1200 \text{ 时,} \\ 0 & \text{其他.} \end{cases} \tag{3.9.16}$$

对于 y 的每一个值我们需要对方程 (3.9.16) 中的函数关于 z 积分来得到 Y 的边际概率密度函数. 看一下使得方程 (3.9.16) 中的函数为正的点对 (y,z) 的集合的图是有帮助的. 图 3-22 中用阴影部分表示该部分区域. 对于 $1800 < y \leqslant 2200$，我们必须对于 z 从 1000 到 $y-800$ 作积分，当 $2200 < y \leqslant 4800$，我们必须对于 z 从 $y-1200$ 到 $y-800$ 作积分，当 $4800 < y < 5200$，我们必须对于 z 从 $y-1200$ 到 4000 作积分. 由于方程 (3.9.16) 中的函数为正时，它是一个常数，积分等于这个常数乘以 z 取值的区间长度. 因此 Y 的概率密度函数是

$$g(y) = \begin{cases} 8.333 \times 10^{-7}(y-1800) & \text{当 } 1800 < y \leqslant 2200 \text{ 时,} \\ 3.333 \times 10^{-4} & \text{当 } 2200 < y \leqslant 4800 \text{ 时,} \\ 8.333 \times 10^{-7}(5200-y) & \text{当 } 4800 < y < 5200 \text{ 时,} \\ 0 & \text{其他.} \end{cases}$$

◀

173

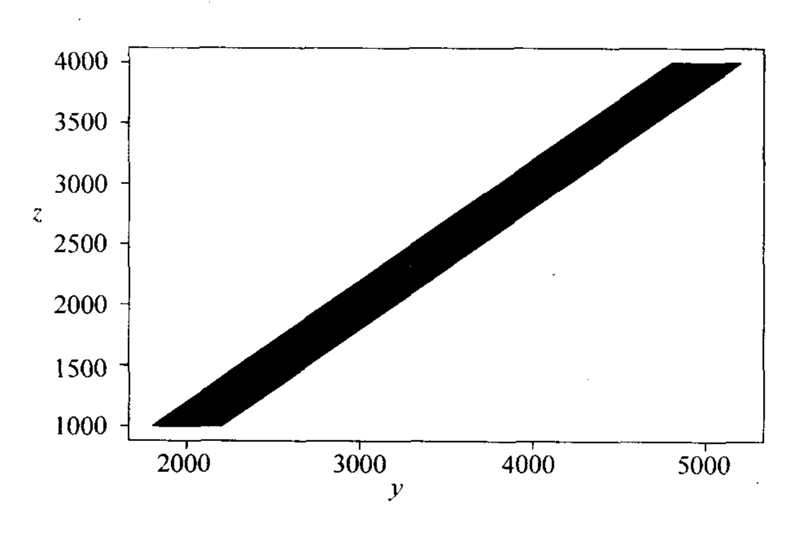


图 3-22　方程 (3.9.16) 中的函数为正的区域

3.9.6　极差

现在设 $X_1, \cdots, X_n$ 是取自概率密度函数为 f、分布函数为 F 的连续随机样本的 n 个

随机变量，和该节的前面相同，定义随机变量 Y_1 和 Y_n 如下：

$$Y_1 = \min\{X_1, \cdots, X_n\} \text{和} \ Y_n = \max\{X_1, \cdots, X_n\}. \tag{3.9.17}$$

随机变量 $W = Y_n - Y_1$ 称为子样极差，换句话说，极差 W 等于样本中的最大值和最小值之差. 下面我们将求 W 的概率密度函数.

样本 Y_1 和 Y_n 的联合概率密度函数 $g(y_1, y_n)$ 由方程(3.9.1)表示. 如果我们令 $Z = Y_1$，那么从 Y_1 和 Y_n 到 W 和 Z 的变换是一个一一对应的线性变换. 其逆变换可用下式表示：

$$Y_1 = Z,$$
$$Y_n = W + Z.$$

对于该变换，$|\boldsymbol{J}| = 1$. 因此，可以通过分别用 z 和 $w+z$ 代替方程(3.9.1)中的 y_1 和 y_n 得到 W 和 Z 的联合概率密度函数 $h(w,z)$. 结果是：

当 $w>0$ 且 $-\infty<z<\infty$ 时

$$h(w,z) = n(n-1)[F(w+z) - F(z)]^{n-2} f(z) f(w+z). \tag{3.9.18}$$

174

对于其他情况 $h(w,z) = 0$.

可以通过下式得到极差 W 的边际概率密度函数 $h_1(w)$：

$$h_1(w) = \int_{-\infty}^{\infty} h(w,z)\,\mathrm{d}z. \tag{3.9.19}$$

例 3.9.5 均匀分布的随机样本的极差 设取自区间 $[0,1]$ 上均匀分布的 n 个随机变量 $X_1, \cdots, X_n$ 组成一个随机样本，我们来求这个随机样本的极差的概率密度函数.

在本例中

$$f(x) = \begin{cases} 1 & \text{当} \ 0<x<1, \\ 0 & \text{其他}. \end{cases}$$

且当 $0<x<1$ 时 $F(x) = x$. 因此，在方程(3.9.18)中，$h(w,z) = 0$，除非 $0<w<1$ 且 $0<z<1-w$. 对于满足这些条件的 w 和 z 的值方程(3.9.18)即为

$$h(w,z) = n(n-1)w^{n-2}.$$

从方程(3.9.19)即得对于 $0<w<1$，W 的概率密度函数是

$$h_1(w) = \int_0^{1-w} n(n-1)w^{n-2}\,\mathrm{d}z = n(n-1)w^{n-2}(1-w).$$

其他情况，$h_1(w) = 0$. ◀

3.9.7 小结

我们把关于一个随机变量的函数的分布的结构推广到关于多个随机变量的多个函数的情形. 如果要得到关于 n 个随机变量的一个函数 r_1 的分布，通常的办法是，首先找到 $n-1$ 个附加的函数 $r_2, \cdots, r_n$ 使得 n 个函数一起组成一个一对一的变换. 再找到 n 个函数的联合概率密度函数，最后，通过对其余 $n-1$ 个变量的积分来得到第一个函数的边际概率密度函数. 该种方法在关于几个随机变量的和和极差的情况阐

述过.

习题

1. 设 X_1 和 X_n 是独立同分布的随机变量，它们都有区间[0,1]上的均匀分布. 求 $Y=X_1+X_n$ 的概率密度函数.
2. 对于习题1的条件，求均值 $(X_1+X_2)/2$ 的概率密度函数.
3. 设三个随机变量 X_1，X_2 和 X_3 有连续型联合分布，其联合概率密度函数如下：

$$f(x_1,x_2,x_3)=\begin{cases}8x_1x_2x_3 & \text{当 } 0<x_i<1(i=1,2,3),\\ 0 & \text{其他.}\end{cases}$$

   且设 $Y_1=X_1$，$Y_2=X_1X_2$ 和 $Y_3=X_1X_2X_3$. 求 Y_1，Y_2 和 Y_3 的联合概率密度函数.
4. 设随机变量 X_1 和 X_2 有连续型联合分布，其联合概率密度函数如下：

$$f(x_1,x_2)=\begin{cases}x_1+x_2 & \text{当 } 0<x_1<1 \text{ 且 } 0<x_2<1,\\ 0 & \text{其他.}\end{cases}$$

   求 $Y=X_1X_2$ 的概率密度函数.

175

5. 设随机变量 X_1 和 X_2 的联合概率密度函数由习题4给出，求 $Z=X_1/X_2$ 的概率密度函数.
6. 设 X 和 Y 是随机变量，它们的联合概率密度函数如下：

$$f(x,y)=\begin{cases}2(x+y) & \text{当 } 0\leqslant x\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

   求 $Z=X+Y$ 的概率密度函数.
7. 设 X_1 和 X_2 是独立同分布的随机变量，它们的概率密度函数如下：

$$f(x)=\begin{cases}e^{-x} & \text{当 } x>0,\\ 0 & \text{其他.}\end{cases}$$

   求 $Y=X_1-X_2$ 的概率密度函数.
8. 设取自区间[0,1]上均匀分布的 n 个随机变量 $X_1,\cdots,X_n$ 组成容量为 n 的随机样本，$Y_n=\max\{X_1,\cdots,X_n\}$. 求满足 $\Pr(Y_n\geqslant 0.99)\geqslant 0.95$ 的 n 的最小值.
9. 设取自区间[0,1]上均匀分布的 n 个随机变量 $X_1,\cdots,X_n$ 组成一个随机样本，随机变量 Y_1 和 Y_n 的定义如方程(3.9.17). 求 $\Pr(Y_1\leqslant 0.1 \text{ 且 } Y_n\leqslant 0.8)$ 的值.
10. 对于习题9的条件，求 $\Pr(Y_1\leqslant 0.1 \text{ 且 } Y_n\geqslant 0.8)$ 的值.
11. 对于习题9的条件，求从 Y_1 到 Y_n 的不包括点1/3的区间的概率.
12. 令 W 表示服从区间[0,1]上均匀分布的 n 个观测值的随机样本的极差，求 $\Pr(W\geqslant 0.9)$ 的值.
13. 求服从区间[-3,5]上均匀分布的 n 个观测值的随机样本的极差的概率密度函数.
14. 设取自区间[0,1]上均匀分布的 n 个随机变量 $X_1,\cdots,X_n$ 组成包含 n 个观测值的随机样本，令 Y 表示第二大的观测值，求 Y 的概率密度函数.

   提示：注意到

$$G(Y)=\Pr(Y\leqslant y)=\Pr(\text{至少有 } n-1 \text{ 个观测值}\leqslant y).$$

先求 Y 的分布函数 G.

15. 证明：如果 $X_1, X_2, \cdots, X_n$ 是独立随机变量，且若 $Y_1 = r_1(X_1), Y_2 = r_2(X_2), \cdots, Y_n = r_n(X_n)$，则 $Y_1, Y_2, \cdots, Y_n$ 也是独立随机变量.
16. 设 $X_1, X_2, \cdots, X_5$ 是五个随机变量，对于所有 $(x_1, x_2, \cdots, x_5) \in R^5$ 的点，它们的联合概率密度函数可以分解为下式：
$$f(x_1, x_2, \cdots, x_5) = g(x_1, x_2)h(x_3, x_4, x_5),$$
此处 g 和 h 都是确定的非负函数. 证明若 $Y_1 = r_1(X_1, X_2)$，$Y_2 = r_2(X_3, X_4, X_5)$，则随机变量 Y_1 和 Y_2 是独立的.
17. 在例 3.9.2 中，用雅可比变换的方法(3.9.4)来证明 Y 和 Z 是独立的，方程(3.9.7)是 Z 的边际概率密度函数.
18. 设给定 Y，X 的条件概率密度函数是 $g_1(x \mid y) = 3x^2/y^3$，当 $0 < x < 1$；其他情况下，为0. Y 的边际概率密度函数是 $f_2(y)$，当 $y \leqslant 0$ 时有 $f_2(y) = 0$，其他情况没有特殊指定. 令 $Z = X/Y$. 证明：Z 和 Y 互相独立，并求 Z 的边际概率密度函数.

176

第4章　数学期望

4.1　单个随机变量的数学期望

随机变量 X 的分布包含了 X 的所有概率信息. 然而，为了表达 X 的这个信息，X 的整个分布往往显得过于繁琐. 当我们想告诉人们关于 X 的一个属性时，并不需要描述其整个分布，而分布的综合特征，例如均值，或者叫作期望值，就变得很有用.

4.1.1　离散分布的数学期望

假设随机变量 X 有离散型分布，其概率函数为 f. X 的数学期望（expectation），记为 $\mathrm{E}(X)$，定义为如下数值：

$$\mathrm{E}(X) = \sum_{\text{所有}x} xf(x). \tag{4.1.1}$$

例 4.1.1　由概率函数计算数学期望　假设随机变量 X 只能取 4 个不同值 -2，0，1，4，且 $\Pr(x=-2)=0.1$，$\Pr(x=0)=0.4$，$\Pr(x=1)=0.3$，$\Pr(x=4)=0.2$. 则

$$\mathrm{E}(X) = -2(0.1)+0(0.4)+1(0.3)+4(0.2)=0.9.$$ ◀

从例 4.1.1 中可以看出，数学期望 $\mathrm{E}(X)$ 并不一定等于 X 的某个可能值.

如果 X 只能取有限多个不同的值，如例 4.1.1，那么等式(4.1.1)中的和式有限. 181 然而，如果 X 的不同可能的取值是一个无穷序列，那么等式(4.1.1)中的和式是一个无穷级数. 给定概率函数，这个级数也许不收敛. 称数学期望 $\mathrm{E}(X)$ 存在，当且仅当等式(4.1.1)中的和式绝对收敛（absolutely convergent），即当且仅当

$$\sum_{\text{所有}x} |x|f(x) < \infty. \tag{4.1.2}$$

换句话说，如果关系式(4.1.2)成立，那么 $\mathrm{E}(X)$ 存在且其值由等式(4.1.1)给出. 如果关系式(4.1.2)不成立，那么 $\mathrm{E}(X)$ 不存在.

注：X 的数学期望只依赖于 X 的分布. 虽然 $\mathrm{E}(X)$ 称为 X 的数学期望，但是它只依赖于 X 的分布. 每两个具有相同分布的随机变量有相同的数学期望，即使它们之间毫无关系. 基于这个原因，我们通常称之为分布的数学期望，即使我们心中不知道具有该分布的随机变量.

4.1.2　连续分布的数学期望

如果随机变量 X 有连续分布，其密度函数为 f，则定义数学期望 $\mathrm{E}(X)$ 如下：

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x. \tag{4.1.3}$$

例 4.1.2 由密度函数计算数学期望 假设具有连续分布的随机变量 X 的密度函数为

$$f(x)=\begin{cases}2x & \text{对于 } 0<x<1,\\ 0 & \text{其他}.\end{cases}$$

则

$$\mathrm{E}(X)=\int_0^1 x(2x)\,\mathrm{d}x=\int_0^1 2x^2\,\mathrm{d}x=\frac{2}{3}.$$

我们也可以说概率密度函数为 f 的分布的数学期望是 2/3. ◀

称连续分布的数学期望 $\mathrm{E}(X)$ 存在，当且仅当等式(4.1.3)中的积分绝对收敛，即当且仅当

$$\int_{-\infty}^{\infty}|x|f(x)\,\mathrm{d}x<\infty.$$

只要 X 是有界的(bounded)随机变量，即存在实数 a 和 b($-\infty<a<b<\infty$)使得 $\Pr(a\leqslant X\leqslant b)=1$，如例 4.1.2，那么 $\mathrm{E}(X)$ 一定存在.

182

4.1.3 数学期望的解释

数 $\mathrm{E}(X)$ 称为 X 的期望值或均值. 数学期望，期望值以及均值可交替使用.

1. 均值与重心的关系

随机变量的数学期望，或者等价地，其分布的均值，可以看作该分布的重心. 为了解释这个概念，例如考虑如图 4-1 所示的概率函数. 可将 x 轴看作一根标示着附有重量的点位置的长轴，如果在轴的每个点 x_j 处有重量 $f(x_j)$，则若在 $\mathrm{E}(X)$ 处有个支点的话，这根棒将会达到平衡.

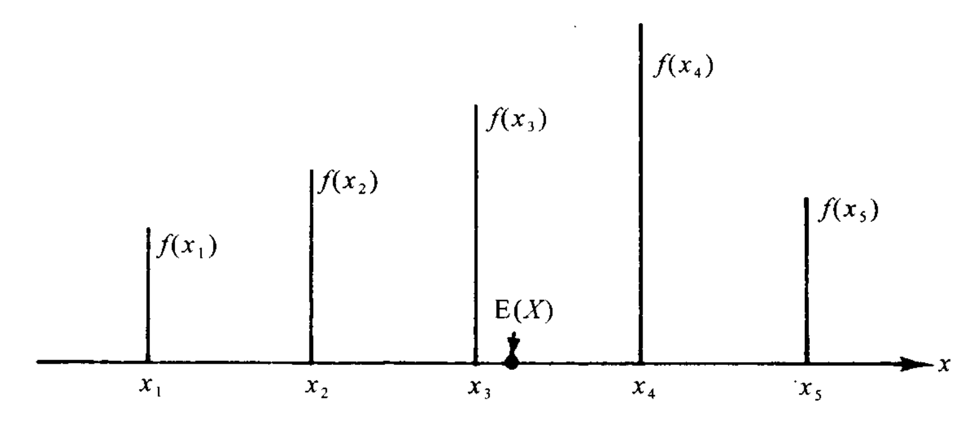
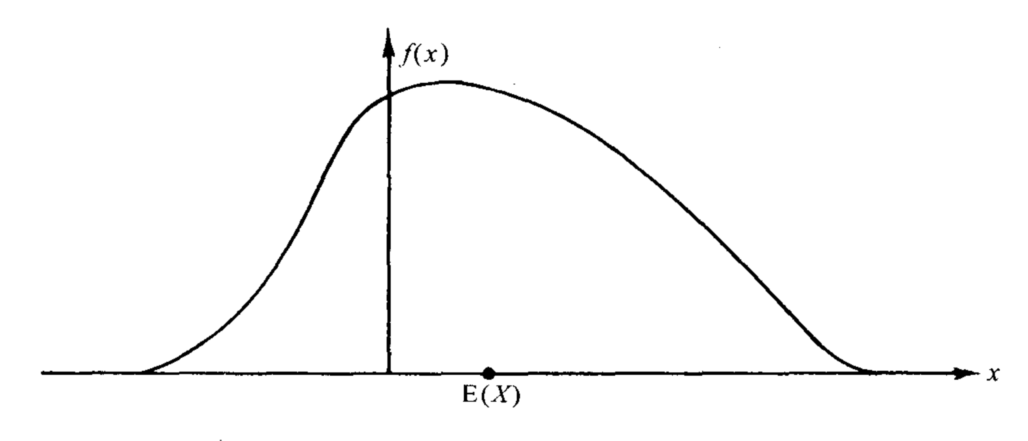


图 4-1 离散分布的均值

现在考虑如图 4-2 所示的密度函数. 此时可将 x 轴看作一根长的其上质量连续变化的棒，如果棒上每点 x 处的密度是 $f(x)$，则这根棒的重心处于点 $\mathrm{E}(X)$，如果在这点有支点的话，该棒将会达到平衡.

从该分布可以看出，x 值很大处的概率值的哪怕一个很小的变化都会对分布的均值产生很大的影响. 例如，如图 4-1 所示的概率函数分布的均值可以被移动到 x 轴上

的任意点，不论该点距原均值点有多远，为此我们只要把某个点 x_j 处减去一个任意小的但是正的概率值，并将它加到距原点充分远的一个点即可.



183

图 4-2 连续分布的均值

现在假设某分布的概率函数或概率密度函数关于 x 轴上某个给定点 x_0 是对称的，换句话说，对 δ 的所有值，有 $f(x_0+\delta)=f(x_0-\delta)$ 成立. 同时假设分布的均值 $E(X)$ 存在，根据均值即重心的解释，则 $E(X)$ 一定等于对称点 x_0. 下面的例子强调了以下事实：在得出 $E(X)=x_0$ 的结论之前必须有均值 $E(X)$ 是存在的作前提.

2. 柯西分布

假设随机变量 X 有连续分布，其密度函数如下：

$$f(x)=\frac{1}{\pi(1+x^2)},\quad -\infty<x<\infty. \tag{4.1.4}$$

称此分布为柯西分布(Cauchy distribution). 我们可以利用以下初等微积分的标准结果验证 $\int_{-\infty}^{\infty}f(x)\,\mathrm{d}x=1$：

$$\frac{\mathrm{d}}{\mathrm{d}x}\tan^{-1}x=\frac{1}{1+x^2},\quad -\infty<x<\infty.$$

等式(4.1.4)表示的密度函数如图 4-3 所示. 该密度函数关于 $x=0$ 对称. 因此，如果柯西分布的均值存在，其值必为 0.

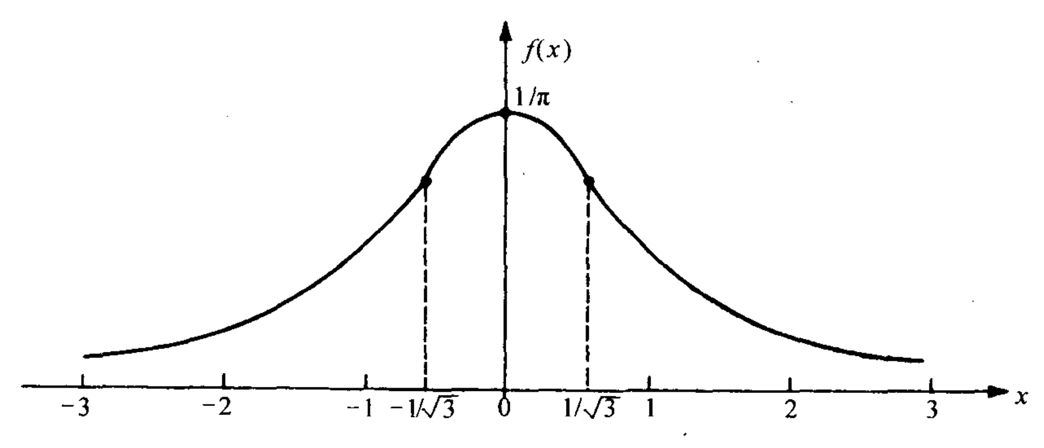


图 4-3 柯西分布的概率密度函数

然而，

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \frac{2}{\pi} \int_{0}^{\infty} \frac{x}{1+x^2} \mathrm{d}x = \infty .$$

因此，柯西分布的均值不存在.

184

柯西分布的均值不存在的原因如下：当曲线 $y=f(x)$ 如图 4-3 时，其尾部足够迅速的趋近于 x 轴使得曲线下部的面积等于 1. 另一方面，如果每一个值 $f(x)$ 乘以 x，则曲线 $y=xf(x)$ 如图 4-4 所示，曲线的尾部很慢地趋近于 x 轴以至于 x 轴与曲线每一部分所夹的面积是无穷的.

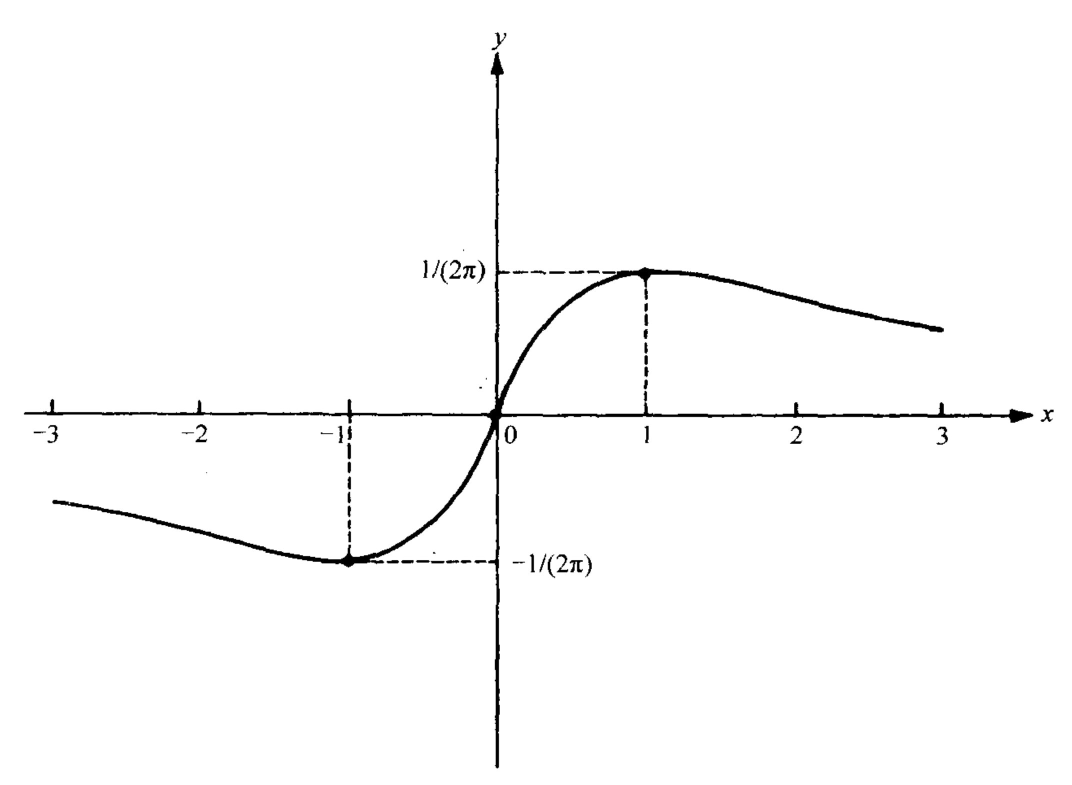


图 4-4 柯西分布的曲线 $y=xf(x)$

4.1.4 函数的数学期望

1. 单个随机变量的函数

如果 X 为随机变量，其密度函数为 f，则函数 $r(X)$ 的数学期望可以通过将数学期望的定义应用于 $r(X)$ 的分布得到：令 $Y=r(X)$，确定 Y 的概率分布，然后应用等式(4.1.1)或(4.1.3)来确定 $\mathrm{E}(Y)$. 例如，假设 Y 有连续分布，密度函数为 g，则

$$\mathrm{E}[r(X)] = \mathrm{E}(Y) = \int_{-\infty}^{\infty} y g(y) \mathrm{d}y. \tag{4.1.5}$$

但是，计算数学期望 $\mathrm{E}[r(X)]$ 时并不一定需要确定 $r(X)$ 的密度函数. 事实上，可以证明 $\mathrm{E}[r(X)]$ 的值可以通过以下关系式来直接计算

185

$$\mathrm{E}[r(X)]=\int_{-\infty}^{\infty} r(x)f(x)\,\mathrm{d}x. \tag{4.1.6}$$

换句话说，可以证明关系式(4.1.5)和(4.1.6)会得出相同的 $\mathrm{E}[r(X)]$值，数学期望 $\mathrm{E}[r(X)]$存在当且仅当

$$\int_{-\infty}^{\infty} |r(x)|f(x)\,\mathrm{d}x<\infty.$$

如果 Y 的分布是离散型，等式(4.1.5)中的积分应换成和式. 如果 X 的分布是离散型，方程(4.1.6)中的积分应换成和式.

本书未给出由关系式(4.1.5)和(4.1.6)推得 $\mathrm{E}[r(X)]$值的等价性的一般证明，而是将用两个特例，来说明这个等价性. 首先，假设 X 的分布是离散型，则 Y 的分布也是离散型. 此时，

$$\begin{aligned}\sum_y yg(y) &= \sum_y y\Pr[r(X)=y]=\sum_y y\sum_{x|r(x)=y} f(x)\\ &=\sum_y\sum_{x|r(x)=y} r(x)f(x)=\sum_x r(x)f(x).\end{aligned}$$

因此，(4.1.5)和(4.1.6)得出相同的值.

其次，假设 X 的分布是连续型，同时如在3.8节中，假设 $r(x)$严格单调递增或严格单调递减，其反函数 $s(y)$可微，则如果在(4.1.6)中将变量 x 换成 $y=r(x)$，

$$\int_{-\infty}^{\infty} r(x)f(x)\,\mathrm{d}x=\int_{-\infty}^{\infty} yf[s(y)]\left|\frac{\mathrm{d}s(y)}{\mathrm{d}y}\right|\mathrm{d}y.$$

由(3.8.3)式可知，等式的右边等于

$$\int_{-\infty}^{\infty} yg(y)\,\mathrm{d}y.$$

因此，(4.1.5)和(4.1.6)得出相同的值.

例4.1.3 确定 $X^{1/2}$的数学期望 假设 X 的概率密度函数由例4.1.2中给出，$Y=X^{1/2}$，则由等式(4.1.6)，

$$\mathrm{E}(Y)=\int_0^1 x^{1/2}(2x)\,\mathrm{d}x=2\int_0^1 x^{3/2}\,\mathrm{d}x=\frac{4}{5}.$$ ◀

注： 一般地，$\mathrm{E}[g(X)]\neq g(\mathrm{E}(X))$. 例4.1.3中，$X^{1/2}$的均值为4/5. 例4.1.2中算得 X 的均值为2/3. 注意 $4/5\neq(2/3)^{1/2}$. 事实上，除非 g 为线性函数，一般$\mathrm{E}[g(X)]\neq g(\mathrm{E}(X))$. 线性函数 g 确实满足 $\mathrm{E}[g(X)]=g(\mathrm{E}(X))$，我们将在定理4.2.1中看到这一点.

例4.1.4 期权定价 假设上市公司 A 的股票价格现在是每股 \$200. 为了激励你为 A 公司工作，你也许会被给予在一年后以 \$200 的价格买入一定数量股票的权利. 如果你认为明年的股价会上涨的话，这个权力就很有价值. 为简化起见，假设一年后的股价 X 是个离散随机变量，只能取两个值(以美元 \$ 计)：260 或 180. 设 $X=260$的概率为 p. 你想计算股票期权的价值，因为或许你想预测卖掉它们的可能性，或许你想比较 A 公司和其他公司的出价. 令 Y 是一年后到期的一股股票的期权的价值. 因为若股价 $X<200$，则没有人愿意花 \$200 去买这支股票，当 $X=180$ 时的

186

期权价值为0. 若 $X=260$，则可以以每股 \$200 买进再立刻以 \$260 卖出，取得每股 \$60 的收益.（为简化起见，忽略股利和买卖股票的交易成本.）则 $Y=h(X)$，其中

$$h(x)=\begin{cases}0 & x=180,\\ 60 & x=260.\end{cases}$$

假设投资者同年可以获得4%的无风险收益.（设4%包含复利.）如果没有其他投资方式，期权的合理成本即一年里 $\mathrm{E}(Y)$ 的现值(present value)，这个值等于 c，满足 $\mathrm{E}(Y)=1.04c$，即一年后的收益应该等于投资者不买此期权获得的价值. 很容易可得 $\mathrm{E}(Y)$：

$$\mathrm{E}(Y)=0\times(1-p)+60\times p=60p.$$

所以，买一股股票的期权的合理价格应为 $c=60p/1.04=57.69p$.

如何确定概率 p 呢？可以通过金融业中的一种标准方法来选择此例中的 p. 该方法假设 X(一年后股票价格)的均值的现值等于当前股价，即买一股股票然后一年后卖出的期望值等于把股票成本无风险的投资一年所得(此例中乘以 1.04). 此例中，意味着 $\mathrm{E}(X)=200\times1.04$. 又 $\mathrm{E}(X)=260p+180(1-p)$，则 $p=0.35$. 一年后 \$200 买一股股票的期权的价格为 \$57.69 ×0.35 = \$20.19. 称此价格为**期权的风险中性价格**(riskneutral price of the option). 可以证明(本节的习题 14)任何不等于 \$20.19 的期权价格会导致市场的不良后果. ◀

2. 多个随机变量的函数

假设给定 n 个随机变量$X_1,\cdots,X_n$的联合密度函数为$f(x_1,\cdots,x_n)$，且 $Y=r(X_1,\cdots,X_n)$. 可以证明由下式可直接确定 $\mathrm{E}(Y)$ 的期望值：

$$\mathrm{E}(Y)=\int\underset{\mathbf{R}^n}{\cdots}\int r(x_1,\cdots,x_n)f(x_1,\cdots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

因此在该种情况下，无需知道 Y 的分布函数即可计算其均值 $\mathrm{E}(Y)$. 187

例4.1.5 确定两个变量的函数的数学期望 假设从方形区域 S 中随机抽取点 (X,Y)，S 包含所有满足 $0\leqslant x\leqslant1$，$0\leqslant y\leqslant1$ 的点(x,y). 要确定 X^2+Y^2 的期望值.

因为 X 和 Y 在区域 S 中是均匀分布的，S 的面积为1，X 和 Y 的联合密度函数为

$$f(x)=\begin{cases}1 & 对于(x,y)\in S,\\ 0 & 其他.\end{cases}$$

因此，

$$\mathrm{E}(X^2+Y^2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^2+y^2)f(x,y)\,\mathrm{d}x\mathrm{d}y=\int_0^1\int_0^1(x^2+y^2)\,\mathrm{d}x\mathrm{d}y=\frac{2}{3}.$$ ◀

4.1.5 小结

一个随机变量的数学期望、期望值或均值，是其分布的一个特征. 如果把概率

分布想象成一根实线上质量的分布，那么均值就是质心．随机变量X的函数r的均值可以由X的分布直接求出，而无需知道$r(X)$的分布．同样的，随机向量X的函数的均值可以由X的分布直接求出．

习题

1. 假设X有区间$[a,b]$上的均匀分布，求X的均值．
2. 从1到100之间随机地抽取一个整数，期望值是多少？
3. 在一个有50个学生的班级里，年龄为i的学生的数量n_i如下表所示：

年龄i	18	19	20	21	25
n_i	20	22	4	3	1

若从班级中随机地选取一名学生，他的年龄的期望值是多少？

4. 假设从THE GIRL PUT ON HER BEAUTIFUL RED HAT这句话中随机地抽取一个单词，X表示所选单词中字母的个数，$\mathrm{E}(X)$的值是多少？
5. 假设从习题4句子中的30个字母中随机抽取一个字母，Y表示抽到字母所在的单词的字母个数，$\mathrm{E}(Y)$的值是多少？
6. 假设随机变量X有连续分布，其概率密度函数由例4.1.2中给出．求$1/X$的数学期望．
7. 假设随机变量X有$[0,1]$上的均匀分布，证明：$1/X$的数学期望不存在．
8. 假设X和Y有连续型联合分布，其连续型密度函数如下：

$$f(x,y)=\begin{cases}12y^2 & 对于\ 0\leqslant y\leqslant x\leqslant 1,\\ 0 & 其他.\end{cases}$$

188

求$\mathrm{E}(XY)$的值．

9. 假设从单位长度的棒上随机选取一点，将棒从该点分为两段．求较长段的长度的期望值．
10. 假设一质点从xy平面的原点出发，往右半平面$x>0$移动，该质点的轨迹是一条直线，与x正半轴的夹角为α，α可正可负，且服从区间$[-\pi/2,\pi/2]$上的均匀分布．令Y为质点轨迹与垂直线$x=1$交点的纵坐标，证明：Y服从柯西分布．
11. 假设随机变量$X_1,\cdots,X_n$构成大小为n的满足区间$[0,1]$上的均匀分布的随机样本．令$Y_1=\min\{X_1,\cdots,X_n\}, Y_n=\max\{X_1,\cdots,X_n\}$，求$\mathrm{E}(Y_1)$和$\mathrm{E}(Y_n)$．
12. 假设随机变量$X_1,\cdots,X_n$构成容量为n的满足分布函数为F的连续分布的随机样本；设随机变量Y_1和Y_n如习题11所定义，求$\mathrm{E}[F(Y_1)]$和$\mathrm{E}[F(Y_n)]$．
13. 一支股票现在每股卖\$110，设一年后的股价为X，X可取值\$100或\$300中的一个．假设你拥有一年后以每股\$150买入的期权，一年期的无风险利率为5.8%．求买入一股的期权的风险中性价格是多少．
14. 考虑例4.1.4中股票期权的定价问题．我们要证明任何不为\$20.19的期权价格都是不合理的．

    a. 假设一个投资者（她已经有一些股票）作如下交易．她又以每股\$200的价格多买入3股股票，以每份\$20.19卖出4份期权．投资者必须以4%的年利率借额外的\$519.24．一年后，投资者必须向买入期权的人以每股\$200的价格卖出4股股票．无论如何，她

必须卖出足够多的股票来偿还年利率4%的借款. 证明：一年后，投资者的净收益(不含误差)与她不作此交易的收益相等，而与股票价格无关. [使得净收益不变的股票和期权的组合称为*无风险投资组合*(risk-free portfolio).]

b. 考虑如a. 中的交易，假设期权价格为\$$x$，$x<20.19$. 证明：投资者的净收益损失了\$$4.16x-84$，且与股票价格无关.

c. 考虑如同a. 中的交易，假设期权价格为\$$x$，$x>20.19$. 证明：投资者的净收益赚得了\$$4.16x-84$，且与股票价格无关.

b. c. 的情形称为*套利机会*(arbitrage opportunities). 在金融市场中的任意长的时间段中这种机会都极少存在. 想象一下如果把3股股票和4份期权变成300万股股票和400万份期权的情形.

15. 在例4.1.4中，我们给出了如何对在将来给定时刻以给定价格买入一股股票的期权进行定价. 这种期权叫作*看涨期权*(call option). 一个*看跌期权*(put option)是在将来给定时刻以给定价格\$$y$卖出一股股票的权利. (如果当你要执行权力时你没有股票,你总可以以市价买入再以\$$y$卖出.)可以将例4.1.4中的原理同样运用到看跌期权的定价上. 考虑例4.1.4中的同种股票，一年后价格为X，分布相同，无风险利率也相同. 求一年后以\$220卖出一股股票的期权的风险中性价格.

4.2 数学期望的性质

在本节中，我们给出一些结果以简化随机变量的某些常用函数的数学期望的计算.

189

4.2.1 基本定理

假设X为一随机变量，其数学期望$E(X)$存在. 我们给出几个关于数学期望基本性质的结果.

定理4.2.1 *如果$Y=aX+b$，其中a和b为常数，则*

$$E(Y)=aE(X)+b.$$

证明： 为方便起见，首先假设X有连续分布，其密度函数为f. 则

$$E(Y)=E(aX+b)=\int_{-\infty}^{\infty}(ax+b)f(x)\,dx$$

$$=a\int_{-\infty}^{\infty}xf(x)\,dx+b\int_{-\infty}^{\infty}f(x)\,dx=aE(X)+b.$$

对于离散分布及更一般分布的证明类似. ■

例4.2.1 计算线性函数的数学期望 假设$E(X)=5$，则

$$E(3X-5)=3E(X)-5=10$$

且

$$E(-3X+15)=-3E(X)+15=0.$$ ◀

由定理4.2.1知，对于任何常数c，$E(c)=c$. 换句话说，如果$X=c$的概率为1，则$E(c)=c$.

例4.2.2 投资 某投资者想从两支股票中作选择作为为期3个月的投资. 一支股票的价格为每股$50，3个月的每股回报率为$R_1$，其中$R_1$为随机变量. 另一支股票的价格为每股$30，3个月的每股回报率为R_2. 投资者的投资总额为$6000. 此例中，假设投资者只买某一种股票.（在例4.2.3中我们考虑投资者购买多于一种股票的策略.）设R_1有$[-10,20]$上的均匀分布，R_2有$[-4.5,10]$上的均匀分布. 我们首先计算投资于每种股票的期望美元价值. 对于第一支股票，$6000可以买120股，回报为$120R_1$，均值为$120E(R_1)=600$.（解4.1节中的例1看为何$E(R_1)=5$.）第二支股票，$6000可以买200股，回报为$200R_2$，均值为$200E(R_2)=550$. 第一支股票有较高的期望回报.

除了计算期望回报，我们也要问这两项投资哪项的风险更大. 现在我们计算每项投资在概率水平0.97下的风险价值(*VaR*).（见例3.3.2.）*VaR*为每项投资回报在$1-0.97=0.03$的分位数的负值. 对第一支股票，回报$120R_1$有区间$[-1200,2400]$上的均匀分布(见3.8节的习题14)，其0.03分位数为(根据例3.3.3)$0.03\times2400+0.97\times(-1200)=-1092$. 因此$VaR=1092$. 对于第二支股票，回报$200R_2$有区间$[-900,2000]$上的均匀分布，其0.03分位数为$0.03\times2000+0.97\times(-900)=-813$. 因此$VaR=813$. 虽然第一支股票有较高的期望回报，但是根据*VaR*，第二支股票却有较低的风险. 在两者之间的选择我们应该怎么来平衡风险和期望收益呢？解决该类问题可参考有关书籍。◀

190

定理4.2.2 *如果存在常数使得$\Pr(X\geqslant a)=1$，则$E(X)\geqslant a$. 如果存在常数使得$\Pr(X\leqslant b)=1$，则$E(X)\leqslant b$.*

证明： 为方便起见，我们仍然假设X有连续分布，其概率密度函数为f，且$\Pr(X\geqslant a)=1$. 则

$$\begin{aligned}E(X)&=\int_{-\infty}^{\infty}xf(x)\,dx\\&=\int_{a}^{\infty}xf(x)\,dx\geqslant\int_{a}^{\infty}af(x)\,dx=a\Pr(X\geqslant a)=a.\end{aligned}$$

定理的另一半证明以及对离散分布或更一般分布的证明类似. ■

由定理4.2.2知，如果$\Pr(a\leqslant X\leqslant b)=1$，则$a\leqslant E(X)\leqslant b$. 也可以证明，如果$\Pr(X\geqslant a)=1$且$E(X)=a$，则$\Pr(X>a)=0$且$\Pr(X=a)=1$.

定理4.2.3 *如果$X_1,\cdots,X_n$为n个随机变量，每一个随机变量的数学期望$E(X_i)$存在$(i=1,\cdots,n)$，则*

$$E(X_1+\cdots+X_n)=E(X_1)+\cdots+E(X_n).$$

证明： 首先假设$n=2$，仍为方便起见，设X_1和X_2有连续型联合分布，联合密度函数为f. 则

$$E(X_1+X_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1+x_2)f(x_1,x_2)\,dx_1dx_2$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1,x_2)\,dx_1dx_2+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2f(x_1,x_2)\,dx_1dx_2$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1,x_2)\,dx_2dx_1+\int_{-\infty}^{\infty}x_2f_2(x_2)\,dx_2$$
$$=\int_{-\infty}^{\infty}x_1f_1(x_1)\,dx_1+\int_{-\infty}^{\infty}x_2f_2(x_2)\,dx_2=E(X_1)+E(X_2)$$

191

其中f_1和f_2分别是X_1和X_2的边际概率密度函数. 对于离散分布或更一般的联合分布的证明类似. 最后利用归纳法可以对每个正整数n证明定理成立. ■

需要强调的是，根据定理4.2.3，几个随机变量的和的数学期望一定总是等于它们各自数学期望的和，而不论这些随机变量是否相互独立.

由定理4.2.1和4.2.3可知，对于所有的常数$a_1,\cdots,a_n$和b，有

$$E(a_1X_1+\cdots+a_nX_n+b)=a_1E(X_1)+\cdots+a_nE(X_n)+b.$$

例4.2.3 投资组合 假设例4.2.2中拥有\$6000的投资者可以同时买两种股票. 设该投资者以每股\$50的价格购买了s_1股第一种股票，以每股\$30的价格购买了s_2股第二种股票. 称这样一种投资的组合为*投资组合*. 在投资总额为\$6000且忽略股数为非整数的情况，则s_1和s_2必须满足

$$50s_1+30s_2=6000,$$

此投资组合的回报为$s_1R_1+s_2R_2$，平均回报为

$$s_1E(R_1)+s_2E(R_2)=5s_1+2.75s_2.$$

例如，取$s_1=54$，$s_2=110$，则平均回报为572.5. ◀

例4.2.4 无放回取样 假设一个盒子里装有红球和蓝球，红球的比例为$p(0\leqslant p\leqslant 1)$. 设从盒子中无放回地随机地选取$n$个球，用$X$表示选取红球的个数. 我们来确定$E(X)$的值.

首先定义n个随机变量$X_1,\cdots,X_n$如下：对于$i=1,\cdots,n$，如果选取的第i个球为红球，记$X_i=1$，如果第i个球为蓝球，记$X_i=0$. 因为这n个球是*无放回地*选取的，则随机变量$X_1,\cdots,X_n$是相互关联的. 然而，可以很容易地得到每一个X_i的边际分布(见1.7节的习题10). 我们可以假设所有的球以随机的顺序放在盒子里，然后选取前n个球. 由于随机性，第i个球为红球的概率简单的为p. 对于$i=1,\cdots,n$，

$$\Pr(X_i=1)=p,\Pr(X_i=0)=1-p. \qquad (4.2.1)$$

因此，$E(X_i)=1(p)+0(1-p)=p$.

由$X_1,\cdots,X_n$的定义可知，$X_1+\cdots+X_n$等于选取到的红球的总数. 因此，$X=X_1+\cdots+X_n$，且由定理4.2.3，

$$E(X)=E(X_1)+\cdots+E(X_n)=np. \qquad (4.2.2)$$

192

◀

注： 只有线性函数 g 满足 $E[g(X)]=g(E(X))$. 定理4.2.1和4.2.3隐含了如果 g 是随机向量 $\boldsymbol{X}$ 的线性函数，则 $E[g(\boldsymbol{X})]=g(E(\boldsymbol{X}))$. 此结果一般只对线性函数 g 成立. 我们已经看到在例4.1.3中 $E[g(\boldsymbol{X})]\neq g(E(\boldsymbol{X}))$.

4.2.2　二项分布的均值

仍然假设一个盒子里装有红球和蓝球，红球的比例为 $p(0\leqslant p\leqslant 1)$. 从盒子中有放回地(with replacement)随机地选取 n 个球. 若用 X 表示样本中红球的个数，则如3.1节中描述，X 有参数为 n 和 p 的二项分布. 现在我们来确定 $E(X)$ 的值.

像前面一样，如果选取的第 i 个球为红球，记 $X_i=1$，否则 $X_i=0$，其中 $i=1,\cdots,n$. 则如前所述，$X=X_1+\cdots+X_n$. 此处，$X_1,\cdots,X_n$ 是互相独立的，每个 X_i 的边际分布由等式(4.2.1)给出. 因此 $E(X_i)=p, i=1,\cdots,n$，由定理4.2.3知

$$E(X)=np. \tag{4.2.3}$$

因此，参数为 n 和 p 的二项分布的均值为 np. 二项分布的概率函数 $f(x)$ 由方程(3.1.2)给出，可以由概率函数直接计算得到它的均值：

$$E(X)=\sum_{x=0}^{n}x\binom{n}{x}p^x q^{n-x}. \tag{4.2.4}$$

因此，由方程(4.2.3)得方程(4.2.4)中和式的值为 np.

由方程(4.2.2)和(4.2.3)可知，不论样本的选取是否可放回，n 个球的样本中红球的期望值都为 np.

例4.2.5　匹配的期望数　假设一个人打印了 n 封信，在 n 个信封上打印了地址，然后随机地把信放入信封里. 令 X 为把信放进正确的信封的数量. 下面求 X 的均值.（在1.7节中我们做了关于此例的更困难的计算.）

如果将第 i 封信放入了正确的信封，记 $X_i=1$，否则记 $X_i=0$，则对于 $i=1,\cdots,n$

$$\Pr(X_i=1)=\frac{1}{n}\text{ 且 }\Pr(X_i=0)=1-\frac{1}{n}.$$

因此

193

$$E(X_i)=\frac{1}{n}\quad\text{对于 } i=1,\cdots,n.$$

因为 $X=X_1+\cdots+X_n$，则

$$E(X)=E(X_1)+\cdots+E(X_n)=\frac{1}{n}+\cdots+\frac{1}{n}=1.$$

因此信和信封正确匹配数的期望值为1，和 n 的值无关. ◀

4.2.3　乘积的数学期望

定理4.2.4　如果 $X_1,\cdots,X_n$ 为 n 个独立的随机变量，每一个的期望 $E(X_i)$ 存在 $(i=1,\cdots,n)$，则

$$\mathrm{E}\left(\prod_{i=1}^{n}X_i\right)=\prod_{i=1}^{n}\mathrm{E}(X_i).$$

证明：仍为方便起见，设 $X_1,\cdots,X_n$ 有连续型联合分布，其联合概率密度函数为f. 我们也记f_i 为 X_i 的边际概率密度函数($i=1,\cdots,n$). 则由于 $X_1,\cdots,X_n$ 是互相独立的，对$(x_1,\cdots,x_n)\in\mathbf{R}^n$ 的每个点，有

$$f(x_1,\cdots,x_n)=\prod_{i=1}^{n}f_i(x_i).$$

因此，

$$\begin{aligned}\mathrm{E}\left(\prod_{i=1}^{n}X_i\right)&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{i=1}^{n}x_i\right)f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\prod_{i=1}^{n}x_if_i(x_i)\right]\mathrm{d}x_1\cdots\mathrm{d}x_n=\prod_{i=1}^{n}\int_{-\infty}^{\infty}x_if_i(x_i)\mathrm{d}x_i=\prod_{i=1}^{n}\mathrm{E}(X_i).\end{aligned}$$

对于离散型分布或更一般分布的证明类似. ■

这里需要强调的是定理4.2.3与定理4.2.4的区别. 如果假设每个数学期望存在，一组随机变量的和的数学期望总是等于其单个数学期望的和. 然而，一组随机变量的积的数学期望并不总是等于各个数学期望的积. 如果随机变量是独立的，那么等式也成立.

例4.2.6 计算随机变量组合的数学期望 若 X_1，X_2，X_3 为独立随机变量，满足 $\mathrm{E}(X_i)=0$和 $\mathrm{E}(X_i^2)=1$，$i=1$，2，3. 计算 $\mathrm{E}[X_1^2(X_2-4X_3)^2]$的值.

194

因为 X_1，X_2，X_3 互相独立，则 X_1^2 和$(X_2-4X_3)^2$ 也独立，从而

$$\begin{aligned}\mathrm{E}[X_1^2(X_2-4X_3)^2]&=\mathrm{E}(X_1^2)\mathrm{E}[(X_2-4X_3)^2]\\&=\mathrm{E}(X_2^2-8X_2X_3+16X_3^2)=\mathrm{E}(X_2^2)-8\mathrm{E}(X_2X_3)+16\mathrm{E}(X_3^2)\\&=1-8\mathrm{E}(X_2)\mathrm{E}(X_3)+16=17.\end{aligned}$$ ◀

4.2.4 扩展研究：非负离散分布的数学期望

数学期望的另一种表示

令 X 为一个只能取值$0,1,2,\cdots$的随机变量，则

$$\mathrm{E}(X)=\sum_{n=0}^{\infty}n\Pr(X=n)=\sum_{n=1}^{\infty}n\Pr(X=n).\tag{4.2.5}$$

考虑下面的概率三角阵列：

$$\begin{array}{cccc}\Pr(X=1)&\Pr(X=2)&\Pr(X=3)&\cdots\\&\Pr(X=2)&\Pr(X=3)&\cdots\\&&\Pr(X=3)&\cdots\\&&&\cdots\end{array}$$

我们有两种不同的方法来计算阵列中的所有元素之和. 第一种，可以将阵列中每一列的元素相加，然后把所有的列相加，由此得到 $\sum_{n=1}^{\infty}n\Pr(X=n)$. 另一种，可以将阵列中每

一行的元素相加，然后把所有的行相加．这样我们可以得到 $\sum_{n=1}^{\infty} \Pr(X \geqslant n)$．因此，

$$\sum_{n=1}^{\infty} n\Pr(X=n) = \sum_{n=1}^{\infty} \Pr(X \geqslant n).$$

由等式(4.2.5)可得

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} \Pr(X \geqslant n). \tag{4.2.6}$$

例4.2.7　试验的期望数　假设一个人重复做一项试验直到成功．设每一次试验成功的概率为 $p(0<p<1)$，则失败的概率为 $q=1-p$，所有的试验都是独立的．如果 X 表示第一次试验成功的试验序数，则可以由下式确定 $\mathrm{E}(X)$：

195

因为总是至少需要一次试验，$\Pr(X \geqslant 1)=1$．当 $n=2,3,\cdots$时，至少需要 n 次试验当且仅当前面 $n-1$ 次试验都失败．因此

$$\Pr(X \geqslant n) = q^{n-1}.$$

由等式(4.2.6)，有

$$\mathrm{E}(X) = 1 + q + q^2 + \cdots = \frac{1}{1-q} = \frac{1}{p}.$$ ◀

4.2.5　小结

随机向量的线性函数的均值为均值的线性函数．特别地，和的均值为均值的和．作为一个例子，参数为 n，p 的二项分布的均值为 np．非线性函数一般没有这样的关系．对于独立的随机变量，乘积的均值为均值的乘积．

习题

1. 设某种股票的回报 R（以每股美元计）有区间$[-3,7]$上的均匀分布，又设每股的成本为$\$1.5$，令 Y 为投资10股的净收益（总回报减去成本）．求 $\mathrm{E}(Y)$．
2. 设三个随机变量 X_1，X_2，X_3，是取自均值为5的分布的随机样本，求 $\mathrm{E}(2X_1-3X_2+X_3-4)$．
3. 设三个随机变量 X_1，X_2，X_3，是取自区间$[0,1]$上的均匀分布的随机样本，求 $\mathrm{E}[(X_1-2X_2+X_3)^2]$．
4. 设随机变量 X 服从区间$[0,1]$上的均匀分布，随机变量 Y 服从区间$[5,9]$上的均匀分布，X 和 Y 互相独立．假设构造一个边长为 X 和 Y 的矩形，求矩形面积的期望值．
5. 设变量 $X_1,\cdots,X_n$ 为取自实轴上密度函数为 f 的连续分布的容量为 n 的随机样本．求样本中观察值落在特定区间$[a,b]$中的个数的期望值．
6. 假设一个质点从实轴的原点出发，在实轴上以一个单位跳跃移动．每次移动是以概率 $p(0 \leqslant p \leqslant 1)$向左，以概率 $1-p$ 向右．求 n 步跳跃后质点位置的期望值．
7. 设某一赌博者在每一场博弈中都以等概率的赢或输．如果赢了，他的赌本翻倍；如果输了，其赌本减半．如果他的初始赌本为 c，问 n 场独立的博弈后其赌本的期望值是多少？
8. 设一个班级里有10个男生，15个女生，从班级中不放回地随机挑选8名学生．设 X 表示选中的男生的个数，Y 为选中的女生的个数．求 $\mathrm{E}(X-Y)$．

196

9. 设在大量产品中不合格品的比例为 p，如在一批产品中随机地选取 n 个样本. 用 X 表示样本中不合格品的个数，用 Y 表示合格品的个数，求 $\mathrm{E}(X-Y)$.
10. 设重复地抛一个均匀硬币直到第一次出现"正面"为止，
    a. 需要抛硬币次数的期望值为多少？
    b. 在第一次出现"正面"之前出现"反面"的期望值为多少？
11. 设重复地抛一枚均匀硬币直到恰好出现 k 次"正面". 求需要抛硬币次数的期望值为多少？(提示：总的抛硬币次数 X 记为 $X_1+\cdots+X_k$，其中 X_i 是在得到 $i-1$ 个"正面"后为了得到第 i 个"正面"需要抛的次数).
12. 设例 4.2.2 和 4.2.3 中的两个回报随机变量 R_1，R_2 是互相独立的. 考虑例 4.2.3 中投资组合，其中第一种股票 $s_1=54$ 股，第二种股票 $s_2=110$.
    a. 证明投资组合的价值变化的密度函数为

$$f(x)=\begin{cases}3.87\times10^{-7}(x+1035) & \text{如果 } -1035<x<560,\\ 6.1728\times10^{-4} & \text{如果 } 560\leqslant x\leqslant 585,\\ 3.87\times10^{-7}(2180-x) & \text{如果 } 585<x<2180,\\ 0 & \text{其他.}\end{cases}$$

    提示：参见例 3.9.4.

    b. 求投资组合在概率水平为 0.97 的风险价值(VaR).

4.3 方　差

尽管分布的均值是一个对于分布性质的有用概括，但是它并没有包含太多关于分布的信息. 例如，一个均值为 2 的随机变量 X 和一个满足 $\Pr(Y=2)=1$ 的常值随机变量 Y 具有相同的均值，即使这里 X 不是常值的. 为了在该种情况下把 X 的分布和 Y 的分布区分开来，给出一个关于 X 的分布的分散程度的度量或许会是一种有效的手段. 方差就是这样一种度量. 标准差是方差的平方根.

4.3.1 方差和标准差的定义

假设 X 是一个具有均值 $\mu=\mathrm{E}(X)$ 的随机变量. X 的方差(variance)，记为 $\mathrm{Var}(X)$，定义为：

$$\mathrm{Var}(X)=\mathrm{E}[(X-\mu)^2]. \tag{4.3.1}$$

因为 $\mathrm{Var}(X)$ 是非负随机变量 $(X-\mu)^2$ 的期望，所以 $\mathrm{Var}(X)\geqslant 0$. 我们必须谨记，在式(4.3.1)中，期望值可能有限也能是无穷大. 如果期望值不是有限的，称 $\mathrm{Var}(X)$ 不存在. 如果 X 的可能取值是有界的，则 $\mathrm{Var}(X)$ 必然存在.

注： *方差仅仅与分布有关.* 随机变量 X 的方差（以及随后马上要定义的标准差）仅仅与 X 的分布有关，就像 X 的期望也仅仅和它的分布有关一样. 事实上，所有能从概率函数或者概率密度函数计算得到的量，都只和分布有关. 两个具有相同分布的随机变量具有相同的方差，它们之间甚至可以毫无关系.

分布的方差提供了一个该分布在其均值μ周围的分散程度的度量. 如果方差值很小，则表示这个概率分布紧密地集中在μ周围；如果方差值很大，则一般表示这个概率分布在μ周围分散得很广. 然而，只要在实轴上离原点足够远的地方放置一个非常小的正值概率，一个分布的方差、以及它的均值就可以变得任意大.

197

随机变量或者分布的*标准差*(standard deviation)定义为方差的非负平方根，一个给定随机变量的标准差一般记为σ，方差一般记为σ^2.

例4.3.1　计算方差和标准差　假设随机变量X能够等概率地取到以下5个值：-2、0、1、3和4，我们来计算X的方差和标准差.

在这个例子里

$$E(X)=\frac{1}{5}(-2+0+1+3+4)=1.2.$$

令$\mu=E(X)=1.2$，并且定义$W=(X-\mu)^2$，则可得$\mathrm{Var}(X)=E(W)$. 我们可以很容易地计算出W的概率函数f

X	-2	0	1	3	4
w	10.24	1.44	0.04	3.24	7.84
$f(w)$	1/5	1/5	1/5	1/5	1/5

从而可以得到：

$$\mathrm{Var}(X)=E(W)=\frac{1}{5}[10.24+1.44+0.04+3.24+7.84]=4.56.$$

X的标准差是方差的平方根，即2.135. ◀

4.3.2　方差的性质

现在我们来提出一些关于方差基本性质的定理，在这些定理里面我们假设所有随机变量的方差都是存在的. 第一个定理表明随机变量X的方差不为0，除非X的整个概率分布集中在一个单点上.

定理4.3.1　$\mathrm{Var}(X)=0$当且仅当存在一个常数c使得$\Pr(X=c)=1$.

证明：首先假设存在常数c使得$\Pr(X=c)=1$. 则有$E(X)=c$，并且$\Pr[(X-c)^2=0]=1$，所以可以得到：

$$\mathrm{Var}(X)=E[(X-c)^2]=0.$$

反过来，假设$\mathrm{Var}(X)=0$，则有$\Pr[(X-\mu)^2\geqslant 0]=1$但是$E[(X-\mu)^2]=0$，所以，根据定理4.2.2的证明后面给出的注释，可以得到

$$\Pr[(X-\mu)^2=0]=1.$$

198

从而，$\Pr(X=c)=1$. ■

定理4.3.2　对于常数a和b，

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X).$$

证明：如果$E(X)=\mu$，则有$E(aX+b)=a\mu+b$，因此

$$\begin{aligned}\mathrm{Var}(aX+b)&=\mathrm{E}[(aX+b-a\mu-b)^2]=\mathrm{E}[(aX-a\mu)^2]\\&=a^2\mathrm{E}[(X-\mu)^2]=a^2\mathrm{Var}(X).\end{aligned}$$ ■

从定理4.3.2可以得到 $\mathrm{Var}(X+b)=\mathrm{Var}(X)$ 对于每个常数 b 成立. 这个结果从直观上很好理解，因为沿着实轴把 X 的分布整个移动 b 个单位以后，会使分布的均值变化 b 个单位，但是该移动不会影响此分布在其均值周围的分散情况.

类似地，从定理4.3.2还可以得到 $\mathrm{Var}(-X)=\mathrm{Var}(X)$. 这个结果也是很容易从直观上理解，因为把整个 X 的分布在实轴上关于原点取对称以后，会得到一个新的分布，这个分布是原来分布关于原点的镜像，均值会从 μ 变到 $-\mu$，但是分布围绕其均值的总的分散情况并不会改变.

例4.3.2　计算线性函数的方差和标准差　考虑和例4.3.1中一样的随机变量 X，X 可以等概率地取到以下5个值：-2、0、1、3和4. 我们来确定 $Y=4X-7$ 的方差和标准差.

在例4.3.1中，我们算得 X 的均值为 $\mu=1.2$，方差为4.56，根据定理4.3.2

$$\mathrm{Var}(Y)=16\mathrm{Var}(X)=72.96.$$

因此，Y 的标准差 σ 为

$$\sigma=(72.96)^{1/2}=8.54.$$ ◀

下一个定理提供了另外一种计算 $\mathrm{Var}(X)$ 的值的方法.

定理4.3.3　对于每个随机变量 X，$\mathrm{Var}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$.

证明：令 $\mathrm{E}(X)=\mu$，则有：

$$\begin{aligned}\mathrm{Var}(X)&=\mathrm{E}[(X-\mu)^2]=\mathrm{E}(X^2-2\mu X+\mu^2)\\&=\mathrm{E}(X^2)-2\mu\mathrm{E}(X)+\mu^2=\mathrm{E}(X^2)-\mu^2.\end{aligned}$$ ■

例4.3.3　计算方差的另外一种方法　再一次考虑例4.3.3中的随机变量 X，X 可以等概率地取到以下5个值：-2、0、1、3和4. 我们将用定理4.3.3来计算 $\mathrm{Var}(X)$. 在例4.3.1中，我们算得 X 的均值为 $\mu=1.2$，为了利用定理4.3.3，我们需要

$$\mathrm{E}(X^2)=\frac{1}{5}[(-2)^2+0^2+1^2+3^2+4^2]=6.$$

199

因为 $\mathrm{E}(X)=1.2$，定理4.3.3表明：

$$\mathrm{Var}(X)=6-(1.2)^2=4.56,$$

这个结果和按照例4.3.1的方法计算得到的结果是一致的. ◀

定理4.3.4　如果 $X_1,\cdots,X_n$ 是互相独立的随机变量，则有：

$$\mathrm{Var}(X_1+\cdots+X_n)=\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n).$$

证明：首先假设 $n=2$，若 $\mathrm{E}(X_1)=\mu_1,\mathrm{E}(X_2)=\mu_2$，则有：

$$\mathrm{E}(X_1+X_2)=\mu_1+\mu_2.$$

因此：

$$\begin{aligned}\mathrm{Var}(X_1+X_2) &= \mathrm{E}[(X_1+X_2-\mu_1-\mu_2)^2] \\ &= \mathrm{E}[(X_1-\mu_1)^2+(X_2-\mu_2)^2+2(X_1-\mu_1)(X_2-\mu_2)] \\ &= \mathrm{Var}(X_1)+\mathrm{Var}(X_2)+2\mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)].\end{aligned}$$

因为 X_1 和 X_2 是互相独立的，所以

$$\mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] = \mathrm{E}(X_1-\mu_1)E(X_2-\mu_2) = (\mu_1-\mu_1)(\mu_2-\mu_2) = 0.$$

从而可以得到：

$$\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_1)+\mathrm{Var}(X_2).$$

至此，根据数学归纳法，可以证明定理对于每个正整数 n 成立. ■

需要强调的是，定理4.3.4中的随机变量必须是相互独立的，将会在4.6节讨论非独立的随机变量的和的方差. 综合定理4.3.2和定理4.3.4，可以得出以下推论：

推论4.3.1 如果 $X_1,\cdots,X_n$ 是互相独立的随机变量，$a_1,\cdots,a_n$ 是任意常数，则有：

$$\mathrm{Var}(a_1X_1+\cdots+a_nX_n+b) = a_1^2\mathrm{Var}(X_1)+\cdots+a_n^2\mathrm{Var}(X_n).$$ ■

例4.3.4 投资组合 一个持有100 000美元的投资者希望构建一个投资组合，该投资组合包含两种可购买的股票中的一种或者两种，有可能的话还包括一些固定利率的投资. 假设这两种股票具有随机回报率：一年期每股回报率分别为 R_1 和 R_2. 假设 R_1 具有均值为6方差为55的分布，同时 R_2 具有均值为4方差为28的分布. 假设第一种股票每股成本为60美元，第二种股票每股成本48美元. 假设资金也能以固定年利率3.6进行投资. 该投资组合包含 s_1 股第一种股票和 s_2 股第二种股票，所有剩下的资金（$\$s_3$）以固定利率投资. 则该投资组合的收益是：

200

$$s_1R_1+s_2R_2+0.036s_3,$$

上式中的系数满足以下限制：

$$60s_1+48s_2+s_3 = 100\,000, \tag{4.3.2}$$

同时 $s_1,s_2,s_3\geqslant 0$. 在这里，我们假设 R_1 和 R_2 是独立的. 投资组合收益的均值和方差分别为：

$$\mathrm{E}(s_1R_1+s_2R_2+0.036s_3) = 6s_1+4s_2+0.036s_3,$$

$$\mathrm{Var}(s_1R_1+s_2R_2+0.036s_3) = 55s_1^2+28s_2^2.$$

一种比较一类投资组合的方法是：如果A的收益的均值至少和B的收益的均值一样大，而且A的方差不比B的大，则称A至少和B一样好.（参考Markowitz, 1987,关于对此类问题的一种经典处理方法.）偏爱较小的方差值的原因在于大的方差值对应的是与均值大的偏离. 对于有相同均值的投资组合，一些关于均值的很大的偏离必定存在比均值小的值，从而导致产生巨大损失的风险. 图4-5是对于在本例中所有可能的投资组合的（均值,方差）的值的描述. 具体来讲，对于每个满足（4.3.2）的（S_1,S_2,S_3），对应于一个位于图4-5中标出的区域中的点. 右下方边界上的点，对于固定的方差来说具有最大的收益均值，对于固定的均值来

说具有最小的方差值. 称这些投资组合为有效的(efficient). 例如, 假设投资者希望得到7000的平均收益. 在图4-5中, 7000以上的垂直线段的纵坐标部分表示可能的所有均值为7000的投资组合的方差值. 图4-5表明, 在这些投资组合中, 具有最小方差的是有效的. 在这个投资组合中 $s_1=524.7$, $s_2=609.7$, $s_3=39\ 250$, 方差为 2.55×10^7. 这样一来, 每个具有大于7000的平均回报的投资组合其方差必然比 2.55×10^7 大; 每个具有小于 2.55×10^7 的方差的投资组合其收益的均值必然比7000小. ◀

201

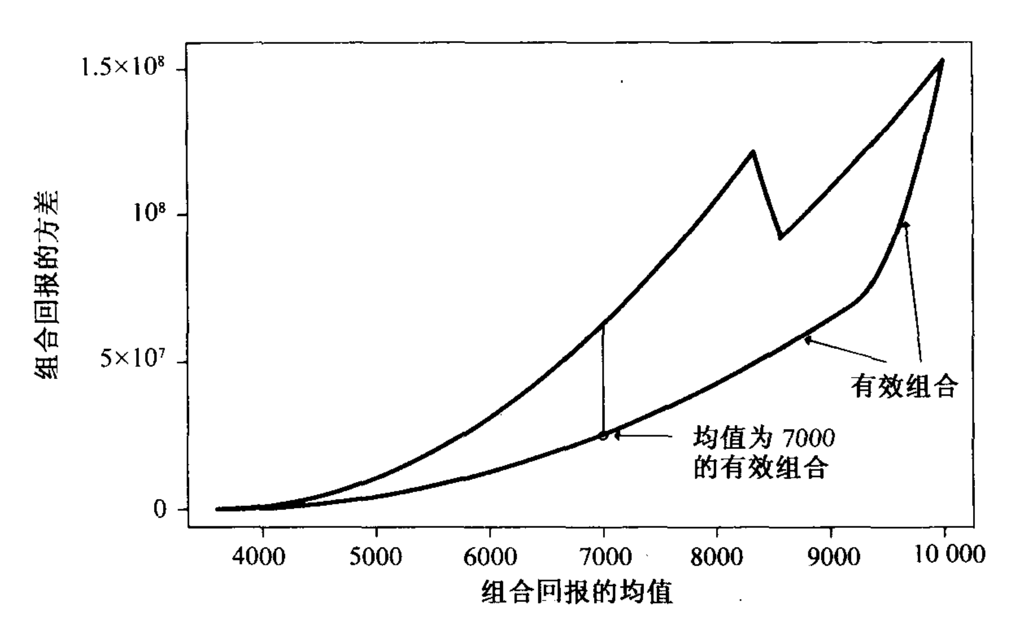


图4-5　投资组合的均值和方差

4.3.3　二项分布的方差

现在我们再次考虑在4.2节中提到的产生一个二项分布的方法. 假设一个箱子装有红色和蓝色的小球, 其中红球的比例为 $p(0\leqslant p\leqslant 1)$. 现从箱子中有放回地随机选取 n 个球. 对于 $i=1,\cdots,n$, 如果第 i 个球为红色, 令 $X_i=1$, 否则令 $X_i=0$. 如果 X 表示取到红球的总次数, 则 $X=X_1+\cdots+X_n$, 而且 X 具有二项分布, 参数为 n 和 p.

因为 $X_1,\cdots,X_n$ 是互相独立的, 从定理4.3.4可以得到:

$$\mathrm{Var}(X)=\sum_{i=1}^{n}\mathrm{Var}(X_i).$$

同时, 对于 $i=1,\cdots,n$

$$\mathrm{E}(X_i)=1(p)+0(1-p)=p$$

并且

$$\mathrm{E}(X_i^2)=1^2(p)+0^2(1-p)=p.$$

因此, 根据定理4.3.3

$$\mathrm{Var}(X_i)=\mathrm{E}(X_i^2)-[\mathrm{E}(X_i)]^2=p-p^2=p(1-p).$$

于是可以得到

$$\mathrm{Var}(X)=np(1-p). \tag{4.3.3}$$

4.3.4 小结

X 的方差，记为 $\mathrm{Var}(X)$，是 $[X-\mathrm{E}(X)]^2$ 的均值，度量了 X 的分布的分散情况. 方差也等于 $\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$. 标准差是方差的平方根. 若 a 和 b 为常数，则 $aX+b$ 的方差为 $a^2\mathrm{Var}(X)$. 独立随机变量的和的方差为他们方差的和. 作为例子，参数为 202 n 和 p 的二项分布的方差为 $np(1-p)$.

习题

1. 设 X 具有区间 $[0,1]$ 上的均匀分布，计算 X 的方差.
2. 假设从以下这个句子里面随机选取一个单词：THE GIRL PUT ON HER BEAUTIFUL RED HAT. 如果用 X 表示选取的单词中含有的字母数，$\mathrm{Var}(X)$ 的值是多少?
3. 对于所有满足 $a<b$ 的数 a 和 b，求在区间 $[a,b]$ 上的均匀分布的方差.
4. 假设 X 是一个随机变量，$\mathrm{E}(X)=\mu$，$\mathrm{Var}(X)=\sigma^2$，证明 $\mathrm{E}[X(X-1)]=\mu(\mu-1)+\sigma^2$.
5. 令 X 为一个随机变量，$\mathrm{E}(X)=\mu$，$\mathrm{Var}(X)=\sigma^2$，并且令 c 为一个任意常数，证明：
$$\mathrm{E}[(X-c)^2]=(\mu-c)^2+\sigma^2.$$
6. 假设 X 和 Y 是互相独立的随机变量，都具有有限的方差，并且 $\mathrm{E}(X)=\mathrm{E}(Y)$，证明：
$$\mathrm{E}[(X-Y)^2]=\mathrm{Var}(X)+\mathrm{Var}(Y).$$
7. 假设 X 和 Y 是互相独立的随机变量，并且 $\mathrm{Var}(X)=\mathrm{Var}(Y)=3$，求以下值：
   a. $\mathrm{Var}(X-Y)$； b. $\mathrm{Var}(2X-3Y+1)$.
8. 构造一个均值存在但是方差不存在的分布的例子.
9. 令 X 具有在整数 $1,\cdots,n$ 上的离散的均匀分布. 计算 X 的方差. 提示：考虑使用公式 $\sum_{k=1}^{n}k^2=n(n+1)\cdot(2n+1)/6$.
10. 考虑例 4.3.4 结尾处的有效投资组合的例子. 假设对于 $i=1$，2，R_i 具有在区间 $[a_i,b_i]$ 上的均匀分布.
    a. 求这两个区间 $[a_1,b_1]$ 和 $[a_2,b_2]$. 提示：区间被均值和方差决定.
    b. 求出例子中投资组合在概率水平 0.97 下的风险价值(VaR). 提示：回顾例 3.9.4 中是怎么求出两个均匀随机变量的和的概率密度函数.

4.4 矩

虽然不像随机变量 X 分布的数字特征那样应用广泛，X 的幂 $X^k(k>2)$ 的均值(称作 X 的矩) 仍具有有用的理论性质. 矩母函数是一个相关的工具，可用它来帮助获得独立随机变量之和的分布以及分布的极限性质.

4.4.1 矩的存在性

对于任意的随机变量 X 和任意正整数 k，称期望 $\mathrm{E}(X^k)$ 为 X 的 k 阶矩. 特别地，

根据这个术语，X 的期望是 X 的一阶矩.

X 的 k 阶矩存在当且仅当 $E(|X^k|)<\infty$. 如果随机变量 X 有界，即，存在常数 a 和 b，满足 $\Pr(a\leqslant X\leqslant b)=1$，则 X 的任意阶矩存在. 然而，当 X 无界时，X 的任意阶矩也有可能存在. 下面的定理表明，当 X 的 k 阶矩存在时，比 k 低阶的矩也存在.

定理 4.4.1 如果对于某个正整数 k，$E(|X^k|)<\infty$ 成立，则 $E(|X^j|)<\infty$，对于任意的满足 $j<k$ 的正整数 j 成立.

203

证明： 不妨假设 X 的分布是连续的，它的概率密度函数是 f，则

$$E(|X|^j)=\int_{-\infty}^{\infty}|x|^jf(x)\,dx=\int_{|x|\leqslant 1}|x|^jf(x)\,dx+\int_{|x|>1}|x|^jf(x)\,dx$$

$$\leqslant\int_{|x|\leqslant 1}1\ f(x)\,dx+\int_{|x|>1}|x|^kf(x)\,dx\leqslant\Pr(|X|\leqslant 1)+E(|X|^k).$$

由假设 $E(|X^k|)<\infty$，因此 $E(|X|^j)<\infty$. 对于离散型和更一般的分布类似可证. ■

特别地，由定理 4.4.1，如果 $E(|X^2|)<\infty$，随机变量 X 的期望和方差均存在.

中心矩

假设随机变量 X，期望是 $E(X)=\mu$. 对于任意正整数 k，称 $E[(X-\mu)^k]$ 为 X 的 k 阶中心矩，或称之为 X 关于均值的 k 阶矩. 特别地，根据该术语，X 的方差是 X 的二阶中心矩.

对于任意分布，一阶中心矩必为 0，因为：

$$E(X-\mu)=\mu-\mu=0.$$

进一步地，如果 X 的分布关于均值 μ 是对称的，如果对于一个给定的奇数 k，中心矩 $E[(X-\mu)^k]$ 存在，则 $E[(X-\mu)^k]$ 的值一定是 0，因为此期望的正部和负部正好相互抵消.

例 4.4.1 对称的概率密度函数 设 X 的分布是连续的，具有如下形式的概率密度函数：

$$f(x)=ce^{-(x-3)^2}\quad -\infty<x<\infty.$$

下面来计算 X 的均值和奇数阶中心矩.

可以证明，对任意正整数 k，

$$\int_{-\infty}^{\infty}|x|^ke^{-(x-3)^2}dx<\infty\quad -\infty<x<\infty.$$

因此，X 的所有阶矩都存在. 进一步地，$f(x)$ 关于点 $x=3$ 对称，所以 $E(X)=3$. 由于这种对称性，对于任意的正奇整数 k，有 $E[(X-3)^k]=0$. ◀

204

4.4.2 矩母函数

考虑给定的随机变量 X，对于任意实数 t，令

$$\psi(t)=E(e^{tX}).\tag{4.4.1}$$

称函数 ψ 为 X 的矩母函数.（简记为 m. g. f）

注： X 的矩母函数仅依赖于 X 的分布．因为矩母函数是 X 的函数的期望值，所以它必定依赖于 X 的分布．如果 X 和 Y 有同样的分布，那么它们一定有相同的矩母函数.

如果随机变量 X 有界，则方程(4.4.1)中的期望对于 t 的所有值都存在．因此，在此例中，X 的矩母函数对于 t 的所有值都存在．另一方面，如果 X 无界，则矩母函数对于 t 的一些值也许存在，对于另一些值也许不存在．然而，从方程(4.4.1)可以看出，对于任意随机变量 X，矩母函数 $\psi(t)$ 在 $t=0$ 处一定存在，其值为 $\psi(0)=\mathrm{E}(1)=1$.

假设对 $t=0$ 点处的某个开区间中 t 的所有值，随机变量 X 的矩母函数都存在，则导数 $\psi'(t)$ 在 $t=0$ 处存在；因此，在 $t=0$ 处，方程(4.4.1)中期望的的导数一定等于导数的期望．（该定理的证明超出了本书的范围）．因此，

$$\psi'(0)=\left[\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}(\mathrm{e}^{tX})\right]_{t=0}=\mathrm{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{tX})_{t=0}\right].$$

但是

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{tX})_{t=0}=(X\mathrm{e}^{tX})_{t=0}=X.$$

从而

$$\psi'(0)=\mathrm{E}(X).$$

也就是说，矩母函数 $\psi(t)$ 在 $t=0$ 处的导数是 X 的均值.

更一般地，如果对于 $t=0$ 点处的某个开区间中 t 的所有的值，随机变量 X 的矩母函数 $\psi(t)$ 都存在，则 X 的所有阶矩 $\mathrm{E}(X^k)$ 都存在 $(k=1,2,\cdots)$．进而，可以证明 $\psi(t)$ 在 $t=0$ 处任意阶可导．对于 $n=1,2,\cdots$ 在 $t=0$ 处，n 阶导数 $\psi^{n}(0)$ 满足以下的关系：

$$\psi^{n}(0)=\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}\mathrm{E}(\mathrm{e}^{tX})\right]_{t=0}=\mathrm{E}\left[\frac{\mathrm{d}^n}{\mathrm{d}t^n}(\mathrm{e}^{tX})_{t=0}\right]=\mathrm{E}[(X^n\mathrm{e}^{tX})_{t=0}]=\mathrm{E}(X^n).$$

因此

205

$$\psi'(0)=\mathrm{E}(X),\psi''(0)=\mathrm{E}(X^2),\psi'''(0)=\mathrm{E}(X^3),\text{等等}.$$

例 4.4.2　计算一个矩母函数　假定随机变量 X 的密度函数是：

$$f(x)=\begin{cases}\mathrm{e}^{-x} & x>0,\\ 0 & x\leqslant 0.\end{cases}$$

下面来计算 X 的矩母函数和 $\mathrm{Var}(X)$.

对于每个实数 t

$$\psi(t)=\mathrm{E}(\mathrm{e}^{tX})=\int_0^\infty \mathrm{e}^{tx}\mathrm{e}^{-x}\,\mathrm{d}x=\int_0^\infty \mathrm{e}^{(t-1)x}\,\mathrm{d}x.$$

该方程的最后一个积分有限当且仅当 $t<1$. 因此，$\psi(t)$ 存在当且仅当 $t<1$. 对于每个这样的 t 值，

$$\psi(t)=\frac{1}{1-t}.$$

因为对于 $t=0$ 处的某个开区间内的所有的 t 值，$\psi(t)$ 都是有限的，X 的所有阶矩都存在. ψ 的前两阶导数分别为：

$$\psi'(t)=\frac{1}{(1-t)^2}\text{且}\ \psi''(t)=\frac{1}{(1-t)^3}.$$

因此

$$E(X)=\psi'(0)=1, E(X^2)=\psi''(0)=2.$$

进而有

$$\mathrm{Var}(X)=\psi''(0)-[\psi'(0)]^2=1.$$

◀

4.4.3 矩母函数的性质

下面介绍关于矩母函数的 3 个基本定理.

定理 4.4.2 令随机变量 X 的矩母函数是 ψ_1，令 $Y=aX+b$，其中 a 和 b 是给定常数，令 ψ_2 表示 Y 的矩母函数. 则对于 t 的每个值，$\psi_1(at)$ 存在.

$$\psi_2(t)=e^{bt}\psi_1(at). \tag{4.4.2}$$

证明： 由矩母函数的定义，

$$\psi_2(t)=E(e^{tY})=E[e^{t(aX+b)}]=e^{bt}E(e^{atX})=e^{bt}\psi_1(at).$$

■

例 4.4.3 计算线性函数的矩母函数 假设 X 的分布由例 4.4.2 给出，则对于 $t<1$，X 的矩母函数是：

$$\psi_1(t)=\frac{1}{1-t}.$$

206

如果 $Y=3-2X$，则 Y 的矩母函数在 $t>-1/2$ 时存在，并且它的值为：

$$\psi_2(t)=e^{3t}\psi_1(-2t)=\frac{e^{3t}}{1+2t}.$$

◀

下一个定理表明，任意个独立随机变量和的矩母函数具有非常简单的形式. 因为具有这种性质，使得矩母函数成为研究这种和的重要工具.

定理 4.4.3 假设 $X_1,\cdots,X_n$ 是 n 个独立的随机变量，对于 $i=1,\cdots,n$，令 ψ_i 表示 X_i 的矩母函数. 令 $Y=X_1+\cdots+X_n$，Y 的矩母函数是 ψ. 则对于任意的 t，ψ_i 存在，$i=1,\cdots,n$

$$\psi(t)=\prod_{i=1}^{n}\psi_i(t). \tag{4.4.3}$$

证明： 由定义，

$$\psi(t)=E(e^{tY})=E[e^{t(X_1+\cdots+X_n)}]=E\left(\prod_{i=1}^{n}e^{tX_i}\right).$$

因为随机变量 $X_1,\cdots,X_n$ 独立，根据定理4.2.4

$$E\left(\prod_{i=1}^{n} e^{tX_i}\right) = \prod_{i=1}^{n} E(e^{tX_i}).$$

因此，

$$\psi(t) = \prod_{i=1}^{n} \psi_i(t). \quad \blacksquare$$

1. 二项分布的矩母函数

假设随机变量 X 是参数为 n，p 的二项分布. 在4.2和4.3节，将 X 视为 n 个随机变量 $X_1,\cdots,X_n$ 的和，求出了 X 的期望和方差. 在这个表示中，每个变量 X_i 的分布是：

$$\Pr(X_i=1)=p \quad \Pr(X_i=0)=q=1-p.$$

现在，用此表示来确定 $X=X_1+\cdots+X_n$ 的矩母函数.

因为每个随机变量 $X_1,\cdots,X_n$ 具有相同的分布，所以每个变量的矩母函数相同. 对于 $i=1,\cdots,n,X_i$ 的矩母函数是

$$\psi_i(t)=E(e^{tX_i})=(e^t)\Pr(X_i=1)+(1)\Pr(X_i=0)=pe^t+q.$$

由定理4.4.3，在这个例子中，X 的矩母函数是：

207

$$\psi(t)=(pe^t+q)^n. \tag{4.4.4}$$

2. 矩母函数的唯一性

现在来介绍矩母函数的一个更重要的性质. 该性质的证明超出了本书的范围.

定理4.4.4 两个随机变量 X_1 和 X_2，在 $t=0$ 处的某个开区间内有相同的矩母函数，则 X_1 和 X_2 的分布函数一定相同. ■

3. 二项分布的可加性

假设 X_1 和 X_2 是独立随机变量，X_1 有参数为 n_1 和 p 的二项分布，X_2 有参数为 n_2 和 p 的二项分布. 这里，两个分布的参数 p 相同，但不需要 $n_1=n_2$. 下面来确定 X_1+X_2的分布.

令 X_i 的矩母函数是 ψ_i，$i=1$，2，则由方程(4.4.4)有

$$\psi_i(t)=(pe^t+q)^{n_i}.$$

如果记 X_1+X_2 的矩母函数为 ψ，由定理4.4.3，

$$\psi(t)=(pe^t+q)^{n_1+n_2}.$$

由方程(4.4.4)可以看出 ψ 是参数为 n_1+n_2 和 p 的二项分布的矩母函数. 因此，由定理4.4.4，X_1+X_2 必定有二项分布，由此，可以得到下边的结果：

如果 X_1 和 X_2 是独立的随机变量，X_i 是参数为 n_i 和 p 的二项分布($i=1,2$)，则 X_1+X_2 有参数为 n_1+n_2 和 p 的二项分布.

4.4.4 小结

如果随机变量的 k 阶矩存在，则对于任意 $j<k$，j 阶矩也存在. X 的矩母函数，$\psi(t)=E(e^{tX})$，如果它对于 t 在0点处某个邻域内存在，则可以用来求解 X 的矩.

$\psi(t)$在 $t=0$ 的 k 阶导数是 $E(X^k)$. 每一对具有相同矩母函数的分布，必是相同的分布，在此意义下，矩母函数描述了分布的特征.

习题

1. 如果 X 有区间$[a,b]$上的均匀分布，则 X 的 5 阶中心矩是多少？
2. 如果 X 有区间$[a,b]$上的均匀分布，写出 X 的偶数阶中心矩公式.
3. 假设 X 是随机变量，满足 $E(X)=1$，$E(X^2)=2$，$E(X^3)=5$. 计算 X 的三阶中心矩.
4. 假设 X 是随机变量，使得 $E(X^2)$存在. (a) 证明：$E(X^2)\geqslant[E(X)]^2$；(b) 证明：$E(X^2)=[E(X)]^2$ 当且仅当存在常数 c，满足 $\Pr(X=c)=1$.
   提示：$\mathrm{Var}(X)\geqslant 0$.
5. 假设 X 是随机变量，有均值μ，方差 σ^2，且 X 的 4 阶矩存在. 证明：$E[(X-\mu)^4]\geqslant\sigma^4$. [208]
6. 如果 X 有区间$[a,b]$上的均匀分布. 试确定 X 的矩母函数.
7. 假设 X 是随机变量，具有如下的矩母函数：
   $$\psi(t)=\frac{1}{4}(3e^t+e^{-t})\quad -\infty<t<\infty.$$
   求 X 的均值和方差.
8. 假设 X 是随机变量，具有如下的矩母函数：
   $$\psi(t)=e^{t^2+3t}\quad -\infty<t<\infty.$$
   求 X 的均值和方差.
9. 假设随机变量 X 的期望是μ，方差是 σ^2，令 $\psi_1(t)$表示 X 的矩母函数，$-\infty<t<\infty$，令 c 是给定的正常数. 且 Y 是随机变量，它的矩母函数是 $\psi_2(t)=e^{c[\psi_1(t)-1]}\quad -\infty<t<\infty$.
   试用 X 的均值和方差来表示 Y 的均值和方差.
10. 假设随机变量 X 和 Y 独立同分布，每个的矩母函数为
   $$\psi(t)=e^{t^2+3t}\quad -\infty<t<\infty.$$
   试求 $Z=2X-3Y+4$ 的矩母函数.
11. 假设 X 是随机变量，具有如下的矩母函数：
   $$\psi(t)=\frac{1}{5}e^t+\frac{2}{5}e^{4t}+\frac{2}{5}e^{8t}\quad -\infty<t<\infty.$$
   试求 X 的概率分布.（提示：这是一个简单的离散分布.）
12. 假设 X 是随机变量，具有如下的矩母函数：
   $$\psi(t)=\frac{1}{6}(4+e^t+e^{-t})\quad -\infty<t<\infty.$$
   试求 X 的概率分布.
13. 假设 X 有柯西分布. 证明其矩母函数 $\psi(t)$仅在 $t=0$ 处存在.
14. 假设 X 的概率密度函数是
   $$f(x)=\begin{cases}x^{-2} & x>1,\\ 0 & x\leqslant 1.\end{cases}$$
   证明：矩母函数 $\psi(t)$在 $t\leqslant 0$ 时存在，在 $t>0$ 时不存在.

4.5 均值和中位数

尽管分布的均值是它中心位置的一个度量，分布的中位数（见3.3）也是它中心位置的一个度量. 本节给出分布的这两种位置归纳之间的一些比较和对照.

4.5.1 中位数

在4.1节提到实轴上的概率分布的均值是该分布的重心. 从此意义上说，可将分布的均值看作分布的中心. 在实轴上还有一点，可能也可以作为分布的中心. 假设有一点 m_0 将总的概率分为相等的两部分，即在 m_0 左边的概率是1/2，在 m_0 右边的概率也是1/2. 对于连续型分布，3.3节介绍的分布的中位数就是这样的数. 如果存在这样的 m_0，可以很自然的称其为分布的中心. 值得注意的是，对于一些离散的分布，不存在将总概率分成完全相等的两部分的点. 进一步地，对于另一些分布，连续型或离散型，也许存在不只一个这样的点. 因此，将要给出的中位数形式上的定义必须能够包括所有的这些可能.

209

对于每个随机变量 X，将 X 分布的中位数定义为这样的一个数 m，满足 $\Pr(X\leqslant m)\geqslant 1/2, \Pr(X\geqslant m)\geqslant 1/2$,

即，中位数是一个点 m，满足以下的两个条件：

第一，如果 m 包含在 m 左边的 X 值中，则

$$\Pr(X\leqslant m)\geqslant \Pr(X>m).$$

第二，如果 m 包含在 m 右边的 X 值中，则

$$\Pr(X\leqslant m)\geqslant \Pr(X<m).$$

根据该定义，每个分布至少有一个中位数，对于一些分布，某个区间中的所有点都是中位数. 如果存在数 m，满足 $\Pr(X<m)=\Pr(X>m)$，即数 m 将全概率精确分成两部分，m 自然就是分布 X 的中位数. 读者可以验证习题1定义的1/2分位数就是中位数.

例4.5.1 离散分布的中位数 假设 X 具有如下的离散分布：

$$\Pr(X=1)=0.1, \Pr(X=2)=0.2,$$
$$\Pr(X=3)=0.3, \Pr(X=4)=0.4.$$

3是该分布的中位数，因为 $\Pr(X\leqslant 3)=0.6$，此值比1/2大，$\Pr(X\geqslant 3)=0.7$，此值同样比1/2大，而3是该分布唯一的中位数. ◀

例4.5.2 中位数不唯一的离散分布 假设 X 具有如下的离散分布：

$$\Pr(X=1)=0.1, \Pr(X=2)=0.4,$$
$$\Pr(X=3)=0.3, \Pr(X=4)=0.2.$$

此处，$\Pr(X\leqslant 2)=1/2$，$\Pr(X\geqslant 3)=1/2$，因此，闭区间 $2\leqslant m\leqslant 3$ 中的每个 m 都是这个分布的中位数. ◀

例 4.5.3 连续分布的中位数 假设 X 有连续分布，有如下的概率密度函数：

$$f(x)=\begin{cases}4x^3 & 0<x<1,\\ 0 & 其他.\end{cases}$$

该分布唯一的中位数 m 满足：

$$\int_0^m 4x^3\mathrm{d}x=\int_m^1 4x^3\mathrm{d}x=\frac{1}{2}.$$

解得 $m=1/2^{1/4}$. ◀

210

例 4.5.4 中位数不唯一的连续分布 假设 X 有连续分布，有如下的概率密度函数：

$$f(x)=\begin{cases}1/2 & 对于 0\leqslant x\leqslant 1,\\ 1 & 对于 2.5\leqslant x\leqslant 3,\\ 0 & 其他.\end{cases}$$

此处，对于闭区间 $1\leqslant m\leqslant 2.5$ 中的每个 m，$\Pr(X\leqslant m)=\Pr(X\geqslant m)=1/2$. 因此，闭区间 $1\leqslant m\leqslant 2.5$ 中的每个 m 都是这个分布的中位数. ◀

4.5.2 均值和中位数的比较

分布的均值和中位数都可以用来表示随机变量的“平均”值. 本节已经描述了均值的一些重要性质，本书的后面将给出均值的更多性质. 尽管如此，出于很多原因，中位数是比均值更有用的关于平均值的度量. 例如，每个分布都存在中位数，但不是每个分布都存在均值. 4.1 节已经提到，由分布的任何部分移除小的但正的概率量，将这个概率量分配给足够大的 x，则分布的均值将变得非常大. 而另一方面，同样的概率改变，中位数却可能不受影响. 将比中位数大的一个 x 值上的任意大的概率重新分配给一个任意大的 x 值上，新分布的中位数和初始分布的中位数是一样的.

例如，某社区的家庭年平均收入是 30 000 美元. 很有可能的是，该社区里只有很少的家庭实际收入有 30 000 美元，而有少数的家庭的收入比 30 000 大得多. （一个极端的例子，假设有 100 个家庭，99 个家庭的收入是 1000 美元，另一个家庭的收入是 2 901 000.）但是如果这些家庭收入的中位数是 30 000 美元，则至少有一半家庭的收入是 30 000 美元或比 30 000 美元多.

现在来考虑两个具体的问题，在这些问题中，必须预测随机变量 X 的值. 在第一个问题中，能够做的最优的预测是均值. 在第二个问题中，最优预测是中位数.

4.5.3 最小均方误差

假设随机变量 X 的均值是 μ，方差是 σ^2. 又假设在一些试验中，观察到 X 的值，但是必须在观察前预测该值. 做预测的一个依据是选择某个数 d，使得误差 $X-d$ 的平方的期望值达到最小. 称 $E[(X-d)^2]$ 为预测 d 的均方误差(mean squared error)(M. S. E.). 现在来确定使均方误差最小的数 d.

对于每个 d 值

211

$$\mathrm{E}[(X-d)^2]=\mathrm{E}(X^2-2dX+d^2)=\mathrm{E}(X^2)-2d\mu+d^2. \tag{4.5.1}$$

方程(4.5.1)的最后表达式是 d 的二次函数. 由初等的微分知识可知，当 $d=\mu$ 时，该函数达到最小值. 因此为了最小化均方误差，X 的预测值应为它的均值 μ. 而当应用这个预测值时，均方误差为 $\mathrm{E}[(X-\mu)^2]=\sigma^2$.

4.5.4 最小平均绝对误差

另一个预测随机变量 X 值的可能的依据是，选择 d，使得 $\mathrm{E}(|X-d|)$ 最小化. 数 $\mathrm{E}(|X-d|)$ 称为预测的平均绝对误差(mean absolute error)(M. A. E.). 下面将证明当平均绝对误差最小化时，d 是分布 X 的中位数.

定理 4.5.1 令 m 是 X 的分布的中位数，令 d 是任意其他值，则

$$\mathrm{E}(|X-m|)\leqslant \mathrm{E}(|X-d|). \tag{4.5.2}$$

而等号成立当且仅当 d 也是 X 的分布的中位数.

证明： 为方便起见，我们假设 X 有连续分布，概率密度函数为 f. 对于任何其他分布类似可证. 首先假设 $d>m$，则

$$\begin{aligned}
\mathrm{E}(|X-d|)-\mathrm{E}(|X-m|) &= \int_{-\infty}^{\infty}(|x-d|-|x-m|)f(x)\mathrm{d}x \\
&= \int_{-\infty}^{m}(d-m)f(x)\mathrm{d}x+\int_{m}^{d}(d+m-2x)f(x)\mathrm{d}x+\int_{d}^{\infty}(m-d)f(x)\mathrm{d}x \\
&\geqslant \int_{-\infty}^{m}(d-m)f(x)\mathrm{d}x+\int_{m}^{d}(m-d)f(x)\mathrm{d}x+\int_{d}^{\infty}(m-d)f(x)\mathrm{d}x \\
&= (d-m)[\Pr(X\leqslant m)-\Pr(X>m)].
\end{aligned} \tag{4.5.3}$$

因为 m 是 X 的分布的中位数，而

$$\Pr(X\leqslant m)\geqslant 1/2\geqslant \Pr(X>m). \tag{4.5.4}$$

因此关系式(4.5.3)中方括号中的差值非负，则有

$$\mathrm{E}(|X-d|)\geqslant \mathrm{E}(|X-m|). \tag{4.5.5}$$

仅当关系式(4.5.3)和(4.5.4)取等号时而关系式(4.5.5)中等号成立. 通过仔细分析可以发现，在这些不等式中，只有当 d 也是分布 X 的中位数时，等号才成立.

对于任意 d，满足 $d<m$ 同理可证. ■

212

例 4.5.5 预测离散随机变量的值 假设随机变量 X 取以下6个值：0、1、2、3、5、7的概率分别为1/6. 现在来分别确定使均方误差和平均绝对误差最小的预测.

在本例中

$$\mathrm{E}(X)=\frac{1}{6}(0+1+2+3+5+7)=3.$$

因此，使均方误差最小的唯一值是 $d=3$.

此外，闭区间 $2\leqslant m\leqslant 3$ 中的每个 m 都是分布的中位数. 所以，使平均绝对误差最小的 d 满足 $2\leqslant d\leqslant 3$，且只对这些 d 值成立. ◀

4.5.5　小结

X 的中位数是任何满足以下关系式 $\Pr(X \leqslant m) \geqslant 1/2$ 且 $\Pr(X \geqslant m) \geqslant 1/2$ 的数 m. 选择 d 来最小化 $\mathrm{E}(|X-d|)$，则 d 一定是 X 的中位数. 选择 d 来最小化 $\mathrm{E}(|X-d|^2)$，则必须选取 $d=\mathrm{E}(X)$.

习题

1. 证明 3.3 节定义的 1/2 分位数是中位数.
2. 假设 X 有离散分布，概率密度函数为：

$$f(x)=\begin{cases} cx & x=1,2,3,4,5,6, \\ 0 & x\neq 1,2,3,4,5,6. \end{cases}$$

求该分布的所有中位数.

3. 假设 X 有连续分布，概率密度函数为：

$$f(x)=\begin{cases} \mathrm{e}^{-x} & x>0, \\ 0 & x\leqslant 0. \end{cases}$$

求该分布的所有中位数.

4. 在一个由 153 户家庭组成的小型社区中，有 k 个小孩($k=0,1,2\cdots$)的家庭数由下表给出，

小孩数	家庭数
0	21
1	40
2	42
3	27
4 或更多	23

确定每户人家小孩数的均值和中位数.（对于均值，假设大于等于 4 个小孩的家庭仅有 4 个小孩，这个假设对中位数有何影响?）

5. 假设 X 的观测值等概率的来自于两个连续分布，这两个分布的概率密度函数分别为 f 和 g. 假设 $f(x)>0$，$0<x<1$；$f(x)=0$，x 取其他值. $g(x)>0$ 当 $2<x<4$ 时，$g(x)=0$，x 取其他值. 试确定：(a) 均值，(b) X 的分布的中位数.
6. 假设 X 有连续分布，概率密度函数为：

$$f(x)=\begin{cases} 2x & \text{对于 } 0<x<1, \\ 0 & \text{其他}. \end{cases}$$

试确定 d 的值，来最小化 (a) $\mathrm{E}[(X-d)^2]$，(b) $\mathrm{E}(|X-d|)$.

7. 假设在某次考试中，某个人的分数 X 在 $[0,1]$ 中取值，X 有连续分布，概率密度函数为：

$$f(x)=\begin{cases} x+\dfrac{1}{2} & \text{对于 } 0\leqslant x\leqslant 1, \\ 0 & \text{其他}. \end{cases}$$

213

试确定 X 的预测，此预测满足最小化（a）均方误差，（b）平均绝对误差.

8. 假设随机变量 X 的分布是关于点 $x=0$ 对称的，且 $\mathrm{E}(X^4)<\infty$. 试证当 $d=0$ 时 $\mathrm{E}[(X-d)^4]$ 达到最小.
9. 假设火灾可能发生在路上的五个点中任意一点. 这些点分布在 -3，-1，0，1，和 2 处（见图 4-6）. 假设这些点是下一个着火点的概率如图 4-6 所示.

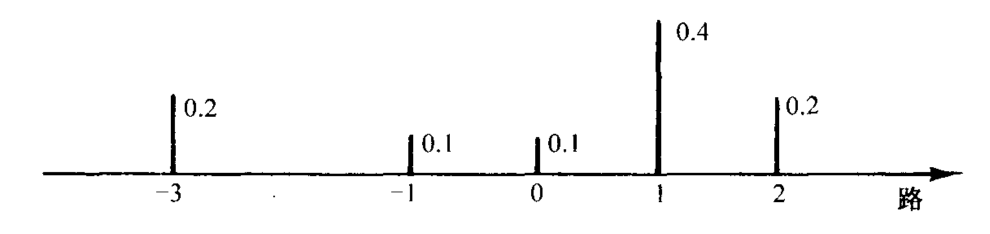


图 4-6　习题 9 的概率

a. 消防车应该在路上的哪个点等候，以便使消防车赶到下一个着火点的距离平方的期望值最小?

b. 消防车应该在路上的哪个点等候，以便使消防车赶到下一个着火点的距离的期望值最小?

10. 如果 n 所房子坐落在一条直路上的不同点，要在这条路上设一个商店，使得它到 n 个房子的距离和最短，商店应该建在哪点?
11. 假设 X 有二项分布，参数为 $n=7$ 和 $p=1/4$. Y 有二项分布，参数为 $n=5$ 和 $p=1/2$. 对于两个随机变量的预测哪一个的均方误差更小?
12. 考虑一个硬币，每次投掷得到头像面的概率是 0.3，假设投掷该硬币 15 次，令 X 表示得到头像的次数.

    a. X 的哪个预测具有最小的均方误差?

    b. X 的哪个预测具有最小的平均绝对误差?
13. 假设 X 的分布关于点 m 对称，证明 m 是 X 的中位数.
14. 求 4.1 节定义的柯西分布的中位数.
15. 假设 X 是随机变量，有分布函数 F. 假设 $a<b$，且 a 和 b 都是 X 的中位数.

    a. 证明：$F(a)=1/2$.

    b. 证明：存在 c 和 d，$c\leq a$ 和 $b\leq d$，满足闭区间 $[c,d]$ 中的每个值都是 X 的中位数.

    c. 假设 X 有离散分布，证明 $F(d)>1/2$.

4.6　协方差和相关系数

当我们对两个随机变量的联合分布感兴趣时，概述两个随机变量相互依赖的程度是很有用的. 协方差和相关系数试图度量这种依赖关系，但他们只捕捉到特殊类型的依赖性，称之为线性依赖性.

4.6.1　协方差

当考虑两个随机变量的联合分布时，随机变量的均值、中位数和方差提供了边际分布的有用信息. 但是，这些数值并没有提供有关两个随机变量之间关系的任何

214

信息，也没有提供它们非独立但同时变化的趋势. 本节和下节将介绍联合分布的一些概述，使得我们能度量两个随机变量关联的程度，确定任意多相关随机变量和的方差，以及通过观察一些相关的其他随机变量来预测一个随机变量的值.

假设 X 和 Y 是随机变量，具有某个给定的联合分布. 令 $E(X)=\mu_X$，$E(Y)=\mu_Y$，$\mathrm{Var}(X)=\sigma_X^2$，$\mathrm{Var}(Y)=\sigma_Y^2$. 则 X 和 Y 的*协方差*，记为 $\mathrm{Cov}(X,Y)$，定义为：

$$\mathrm{Cov}(X,Y)=\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]. \tag{4.6.1}$$

可以证明（见本节最后的习题2），当 $\sigma_X^2<\infty$ 和 $\sigma_Y^2<\infty$ 时，方程(4.6.1)的期望存在，并且 $\mathrm{Cov}(X,Y)$ 有限. 然而，$\mathrm{Cov}(X,Y)$ 的值可以是正、负或0.

例 4.6.1　计算协方差　假设随机变量 X 和 Y 有连续分布，具有如下的联合密度函数

$$f(x,y)=\begin{cases}x+y & \text{对于 } 0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

下面来计算协方差 $\mathrm{Cov}(X,Y)$. 首先必须分别计算 X 和 Y 的期望 μ_X 和 μ_Y. X 的边际概率密度函数是 $f_1(x)=x+1/2$，对于 $0\leqslant x\leqslant 1$. 可以很容易地计算出期望 $\mathrm{E}(X)=7/12$. 由概率密度函数中 X 和 Y 的对称性可见，X 和 Y 有相同的边际分布，$\mathrm{E}(Y)=7/12$. 通过4.1节计算若干个随机变量的函数的均值的方法，也可以计算出协方差，特别地，计算积分

$$\int_0^1\int_0^1\left(x-\frac{7}{12}\right)\left(y-\frac{7}{12}\right)(x+y)\,\mathrm{d}x\,\mathrm{d}y.$$

可直接计算这个积分，虽有些乏味，结果是 $\mathrm{Cov}(X,Y)=-1/144$. ◀

4.6.2 相关系数

X 和 Y 的协方差是用来度量 X 和 Y 同时增大，或一个增大一个减少的程度. 例如，假设 $\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]$ 是正的，则 $X>\mu_X$ 和 $Y>\mu_Y$ 必同时发生，或者 $X<\mu_X$ 和 $Y<\mu_Y$ 同时发生的程度要比 $X>\mu_X$ 和 $Y<\mu_Y$ 同时发生、或者 $X<\mu_X$ 和 $Y>\mu_Y$ 同时发生的程度高. 否则，均值将为负的了. 但是 $\mathrm{Cov}(X,Y)$ 的大小受到 X 和 Y 总的大小的影响. 例如，本节习题5，可以证明 $\mathrm{Cov}(2X,Y)=2\mathrm{Cov}(X,Y)$. 为了获得 X 和 Y 之间相关关系的度量，而不受到一个或另一个随机变量尺度改变的影响，下面定义一个稍稍不同的数量值.

215

如果 $0<\sigma_X^2<\infty$ 且 $0<\sigma_Y^2<\infty$，则 X 和 Y 的*相关系数*(correlation)，记为 $\rho(X,Y)$，定义为：

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}. \tag{4.6.2}$$

为了确定相关系数 $\rho(X,Y)$ 取值的可能范围，需要如下的结果：

定理 4.6.1　施瓦茨不等式　*对于所有的随机变量 U 和 V，*

$$[E(UV)]^2 \leqslant E(U^2)E(V^2). \tag{4.6.3}$$

另外，如果不等式(4.6.3)的右边有限，则不等式(4.6.3)的两端相等当且仅当存在常数 a 和 b，使得 $aU+bV=0$ 的概率是1.

证明： 如果 $E(U^2)=0$，则 $\Pr(U=0)=1$. 因此，$\Pr(UV=0)=1$ 也成立. 所以 $E(UV)=0$，关系式(4.6.3)成立. 同理，如果 $E(V^2)=0$，关系式(4.6.3)也成立. 因此可以假设 $E(U^2)>0$，$E(V^2)>0$. 而如果 $E(U^2)$ 或 $E(V^2)$ 之一是无穷，则关系式(4.6.3)右端也是无穷. 在该种情况下，关系式(4.6.3)必定成立.

下面将证明，假设 $0<E(U^2)<\infty$ 且 $0<E(V^2)<\infty$，对于所有的数字 a 和 b，

$$0 \leqslant E[(aU+bV)^2] = a^2E(U^2)+b^2E(V^2)+2abE(UV) \tag{4.6.4}$$

$$0 \leqslant E[(aU-bV)^2] = a^2E(U^2)+b^2E(V^2)-2abE(UV). \tag{4.6.5}$$

如果令 $a=[E(V)^2]^{1/2}$，$b=[E(U)^2]^{1/2}$，则由关系式(4.6.4)有

$$E(UV) \geqslant -[E(U^2)E(V^2)]^{1/2}. \tag{4.6.6}$$

由关系式(4.6.5)有

$$E(UV) \leqslant [E(U^2)E(V^2)]^{1/2}. \tag{4.6.7}$$

以上两式说明关系式(4.6.3)成立.

最后，假设不等式(4.6.3)的右端有限. 式(4.6.3)两端相等当且仅当式(4.6.6)两端相等，或式(4.6.7)两端相等. 式(4.6.6)两端相等当且仅当(4.6.4)最右边的表达式为0. 依次地，式(4.6.6)两端相等也当且仅当 $E[(aU+bV)^2]=0$，即当且仅当 $aU+bV=0$ 的概率是1. 读者可以检验方程(4.6.7)两端值相等当且仅当 $aU-bV=0$ 的概率是1. ■

如果令 $U=X-\mu_X, V=Y-\mu_Y$，则由定理4.6.1有

216

$$[\mathrm{Cov}(X,Y)]^2 \leqslant \sigma_X^2\sigma_Y^2.$$

依次地，由方程(4.6.2)有 $[\rho(X,Y)]^2 \leqslant 1$，或等价地

$$-1 \leqslant \rho(X,Y) \leqslant 1.$$

如果 $\rho(X,Y)>0$，称 X 和 Y 正相关；如果 $\rho(X,Y)<0$，称 X 和 Y 负相关；如果 $\rho(X,Y)=0$，称 X 和 Y 不相关. 从方程(4.6.2)可以看出，$\mathrm{Cov}(X,Y)$ 和 $\rho(X,Y)$ 必定有相同的符号；即都是正的，或都是负的，或都是0.

例4.6.2　计算相关系数　对于例4.6.1所示的两个随机变量，可以计算它们的相关系数 $\rho(X,Y)$. X，Y 的方差都是11/144. 所以，相关系数 $\rho(X,Y)=-1/11$. ◀

4.6.3　协方差和相关系数的性质

下面来介绍描述协方差和相关系数基本性质的五个定理. 第一个定理给出了计算 $\mathrm{Cov}(X,Y)$ 的另一种方法.

定理4.6.2　对于所有满足 $\sigma_X^2<\infty$ 和 $\sigma_Y^2<\infty$ 的随机变量 X 和 Y，有

$$\mathrm{Cov}(X,Y) = E(XY)-E(X)E(Y). \tag{4.6.8}$$

证明： 根据方程(4.6.1)有

$$\mathrm{Cov}(X,Y)=\mathrm{E}(XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y)=\mathrm{E}(XY)-\mu_X\mathrm{E}(Y)-\mu_Y\mathrm{E}(X)+\mu_X\mu_Y.$$

因为 $\mathrm{E}(X)=\mu_X$，$\mathrm{E}(Y)=\mu_Y$，即得到方程(4.6.8). ■

下一个结果证明互相独立的随机变量必是互不相关的.

定理4.6.3 如果 X 和 Y 是独立的随机变量，满足 $0<\sigma_X^2<\infty$ 且 $0<\sigma_Y^2<\infty$，则

$$\mathrm{Cov}(X,Y)=\rho(X,Y)=0.$$

证明： 如果 X 和 Y 是独立的，则 $\mathrm{E}(XY)=\mathrm{E}(X)\mathrm{E}(Y)$. 因此，由(4.6.8)有，$\mathrm{Cov}(X,Y)=0$. 同理，有 $\rho(X,Y)=0$. ■

一般来说，定理4.6.3的逆命题不成立. 两个非独立的随机变量可以不相关. 事实上，即使 Y 是 X 的函数，也有可能 $\rho(X,Y)=0$，见下例.

例4.6.3 非独立但不相关的随机变量 假设随机变量 X 只取到三个值 -1、0、1，取到每个值的概率相等. 令随机变量 Y 满足 $Y=X^2$. 下面来证明 X 和 Y 非独立但不相关.

在本例中，X 和 Y 显然非独立，因为 Y 的值完全由 X 的值决定. 然而

$$\mathrm{E}(XY)=\mathrm{E}(X^3)=\mathrm{E}(X)=0.$$

217

因为 $\mathrm{E}(XY)=0$ 且 $\mathrm{E}(X)\mathrm{E}(Y)=0$，由定理4.6.2有 $\mathrm{Cov}(X,Y)=0$，即 X 和 Y 不相关. ◀

下一个定理表明如果 Y 是 X 的线性函数，则 X 和 Y 一定相关，事实上，$|\rho(X,Y)|=1$.

定理4.6.4 假设 X 是随机变量，满足 $0<\sigma_X^2<\infty$，令 $Y=aX+b$，其中 a，b 是常数，$a\neq0$. 如果 $a>0$，则 $\rho(X,Y)=1$. 如果 $a<0$，则 $\rho(X,Y)=-1$.

证明： 如果 $Y=aX+b$，则 $\mu_Y=a\mu_X+b$，并且 $Y-\mu_Y=a(X-\mu_X)$. 因此由方程(4.6.1)有

$$\mathrm{Cov}(X,Y)=a\mathrm{E}[(X-\mu_X)^2]=a\sigma_X^2.$$

因为 $\sigma_Y=|a|\sigma_X$，由方程(4.6.2)，定理得证. ■

定理4.6.4有逆命题，即 $|\rho(X,Y)|=1$ 时，X 和 Y 线性相关.（参看习题17）一般地，$\rho(X,Y)$ 的值度量了两个随机变量 X 和 Y 线性相关的程度. 如果 X 和 Y 的联合分布相对集中于 xy 平面的一条直线附近，并且此直线的斜率为正，则典型的，$\rho(X,Y)$ 会相当接近于1. 如果 X 和 Y 的联合分布相对集中于 xy 平面的一条直线附近，并且此直线的斜率为负，则典型的，$\rho(X,Y)$ 会相当接近于 -1. 我们不在这里进一步讨论这些概念，而到5.10节介绍和学习二元正态分布时，将再次讨论这些概念.

注： 相关系数仅仅度量线性关系. $|\rho(X,Y)|$ 的值很大意味着 X 和 Y 线性相关的互相接近，因此是密切相关的. 但 $|\rho(X,Y)|$ 的很小值也不意味着 X 和 Y 相关的不紧密. 事实上，例4.6.3表明了随机变量函数相关，但是相关系数为0.

现在来决定随机变量和的方差，而这些随机变量不一定独立.

定理4.6.5 如果 X 和 Y 是随机变量，满足 $\mathrm{Var}(X)<\infty$ 且 $\mathrm{Var}(Y)<\infty$，则

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y). \tag{4.6.9}$$

证明： 因为 $\mathrm{E}(X+Y)=\mu_X+\mu_Y$，则

$$\begin{aligned}\mathrm{Var}(X+Y)&=\mathrm{E}[(X+Y-\mu_X-\mu_Y)^2]\\&=\mathrm{E}[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)]\\&=\mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y).\end{aligned}$$ ■

218

对于所有的常数 a 和 b，可以证明 $\mathrm{Cov}(aX,bY)=ab\,\mathrm{Cov}(X,Y)$（见本节的习题5）．因此由定理4.6.5，有

$$\mathrm{Var}(aX+bY+c)=a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)+2ab\,\mathrm{Cov}(X,Y). \tag{4.6.10}$$

特别地，

$$\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y). \tag{4.6.11}$$

例4.6.4　投资组合　再次考虑本书例4.3.4的投资者，尝试用＄100 000选择一个组合来投资．对于两种股票的回报采用同样的假设，但是假设两种股票的回报 R_1 和 R_2 的相关系数为 -0.3，表示这两种股票对于一般市场波动的反映趋于相反．第一种股票有 s_1 股、第二种股票有 s_2 股、第三种美元的投资为3.6%的投资组合的协方差是：

$$\mathrm{Var}(s_1R_1+s_2R_2+0.036s_3)=55s_1^2+28s_2^2-0.3\sqrt{55\times28}s_1s_2.$$

下面接着假设(4.3.2)成立．图4-7显示了本例子和例4.3.4中，有效资产组合的均值和方差的关系．注意本例的方差比例4.3.4小得多，这是基于这样的事实，负相关降低了具有正系数的线性组合的方差．◀

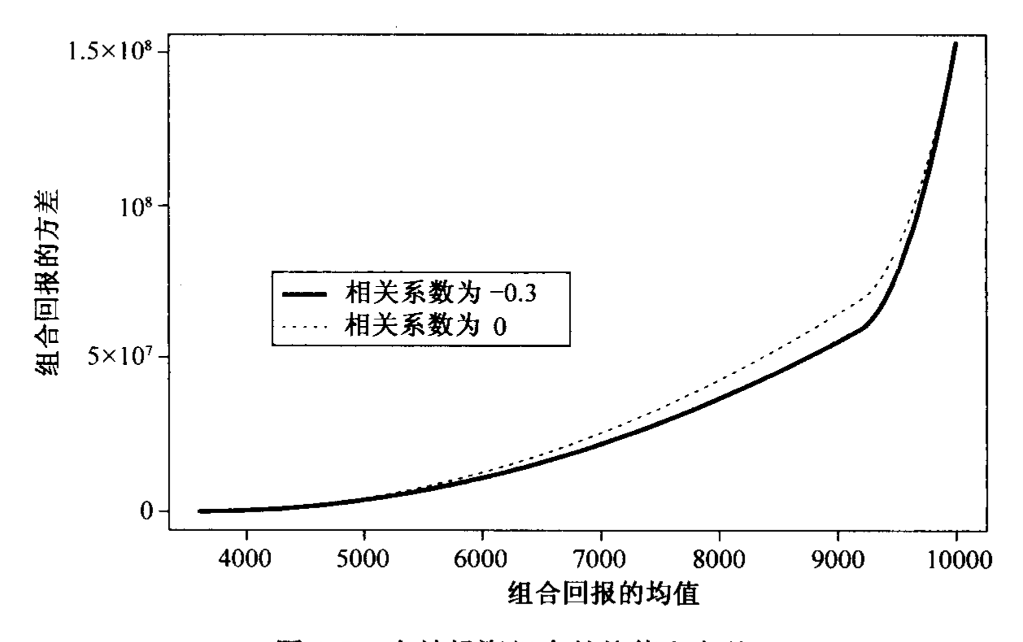


图4-7　有效投资组合的均值和方差

219

可以很容易地将定理4.6.5推广到 n 个随机变量和的方差，具体如下：

定理4.6.6　如果 $X_1,\cdots,X_n$ 是随机变量，满足 $\mathrm{Var}(X_i)<\infty,i=1,\cdots,n$，则

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2\sum_{i<j}\sum \operatorname{Cov}(X_i, X_j). \tag{4.6.12}$$

证明： 对于每个随机变量 Y，$\operatorname{Cov}(Y,Y) = \operatorname{Var}(Y)$. 因此，通过应用本节习题 8 的结果，可以得到下面的关系式：

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \operatorname{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j).$$

将此关系式最后的和式拆成两项：（i）$i=j$ 的项的和，（ii）$i \neq j$ 的项的和. 利用 $\operatorname{Cov}(X_i,X_j) = \operatorname{Cov}(X_j,X_i)$ 的事实，可以得到关系式：

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + \sum_{i\neq j}\sum \operatorname{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2\sum_{i<j}\sum \operatorname{Cov}(X_i, X_j).\end{aligned}$$ ■

定理 4.3.4 表明，如果 $X_1, \cdots, X_n$ 是独立的随机变量，则

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i). \tag{4.6.13}$$

可将该结果推广为：

如果 $X_1, \cdots, X_n$ 是不相关的随机变量（即 $i \neq j$ 时，X_i 和 X_j 不相关），则等式 (4.6.13) 成立.

注： *仅对不相关的随机变量增加方差.* 方程 (4.6.13) 仅对于不相关的随机变量成立. 一般地，要通过定理 4.6.6 来求得随机变量和的方差.

4.6.4 小结

X 和 Y 的协方差为 $\operatorname{Cov}(X,Y) = E\{[X-E(X)][Y-E(Y)]\}$. 相关系数为 $\rho(X,Y) = \operatorname{Cov}(X,Y)/[\operatorname{Var}(X)\operatorname{Var}(Y)]^{1/2}$，该值度量了 X 和 Y 线性相关的程度. 事实上，X 和 Y 完全线性相关当且仅当 $|\rho(X,Y)| = 1$. 随机变量和的方差可以表示为各个随机变量方差的和加上协方差的 2 倍和. 线性函数的方差为

$$\operatorname{Var}(aX + bY + c) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y) + 2ab\operatorname{Cov}(X,Y).$$

[220]

习题

1. 假设 (X,Y) 均匀地分布在半径为 1 的圆内部. 计算 $\operatorname{Cov}(X,Y)$.
2. 证明：如果 $\operatorname{Var}(X) < \infty$ 且 $\operatorname{Var}(Y) < \infty$，则 $\operatorname{Cov}(X,Y)$ 有限. *提示：考虑关系式* $[(X-\mu_X) \pm (Y-\mu_Y)]^2 \geq 0$，证明 $|(X-\mu_X)(Y-\mu_Y)| \leq \frac{1}{2}[(X-\mu_X)^2 + (Y-\mu_Y)^2]$.
3. 假设 X 有区间 $[-2,2]$ 上的均匀分布，且 $Y = X^2$，证明：X 和 Y 不相关.
4. 假设随机变量 X 的分布关于 $x=0$ 对称，$0 < E(X^4) < \infty$，且 $Y = X^2$，证明：X 和 Y 不相关.

5. 对于任意的随机变量 X 和 Y 和任意的常数 a, b, c, d, 证明: $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$.
6. 设 X 和 Y 是随机变量, 满足 $0<\sigma_X^2<\infty$ 和 $0<\sigma_Y^2<\infty$. 假设 $U=aX+b$, $V=cY+d$, 其中 $a\neq 0$, $c\neq 0$. 证明: $\rho(U,V)=\rho(X,Y)$, 如果 $ac>0$; $\rho(U,V)=-\rho(X,Y)$, 如果 $ac<0$.
7. 令 X, Y, Z 是三个随机变量, 使得 $\mathrm{Cov}(X,Z)$ 和 $\mathrm{Cov}(Y,Z)$ 存在, 令 a, b, c 是任意给定的常数. 证明: $\mathrm{Cov}(aX+bY+c,Z)=a\,\mathrm{Cov}(X,Z)+b\,\mathrm{Cov}(Y,Z)$.
8. 假设 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 是随机变量, 使得 $\mathrm{Cov}(X_i,Y_j)$ 存在, $i=1,\cdots,m, j=1,\cdots,n$; 假设 $a_1,\cdots,a_m$ 和 $b_1,\cdots,b_n$ 是常数. 证明: $\mathrm{Cov}\left(\sum_{i=1}^{m} a_iX_i, \sum_{j=1}^{n} b_jY_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j\mathrm{Cov}(X_i,Y_j)$.
9. 假设 X 和 Y 是两个随机变量, 也许非独立, 并且 $\mathrm{Var}(X)=\mathrm{Var}(Y)$. 假设 $0<\mathrm{Var}(X+Y)<\infty$, $0<\mathrm{Var}(X-Y)<\infty$. 证明: 随机变量 $X+Y$ 和 $X-Y$ 不相关.
10. 假设 X 和 Y 是负相关的, $\mathrm{Var}(X+Y)$ 的值比 $\mathrm{Var}(X-Y)$ 大还是小?
11. 证明: 两个随机变量 X 和 Y 不可能同时具有下面的性质: $\mathrm{E}(X)=3$, $\mathrm{E}(Y)=2$, $\mathrm{E}(X^2)=10$, $\mathrm{E}(Y^2)=29$, $\mathrm{E}(XY)=0$.
12. 假设 X, Y 有连续的联合分布, 其联合密度函数如下:

$$f(x,y)=\begin{cases}\frac{1}{3}(x+y) & \text{对于 } 0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1,\\ 0 & \text{其他.}\end{cases}$$

试确定 $\mathrm{Var}(2X-3Y+8)$ 的值.
13. 假设 X 和 Y 是随机变量, 满足 $\mathrm{Var}(X)=9$, $\mathrm{Var}(Y)=4$, $\rho(X,Y)=-1/6$. 试确定 (a) $\mathrm{Var}(X+Y)$, (b) $\mathrm{Var}(X-3Y+4)$.
14. 假设 X、Y 和 Z 是三个随机变量, 满足 $\mathrm{Var}(X)=1$, $\mathrm{Var}(Y)=4$, $\mathrm{Var}(Z)=8$, $\mathrm{Cov}(X,Y)=1$, $\mathrm{Cov}(X,Z)=-1$, $\mathrm{Cov}(Y,Z)=2$. 试确定 (a) $\mathrm{Var}(X+Y+Z)$, (b) $\mathrm{Var}(3X-Y-2Z+1)$.
15. 假设 $X_1,\cdots,X_n$ 是随机变量, 每个随机变量的方差都是 1, 每一对随机变量间的相关系数是 1/4. 试确定 $\mathrm{Var}(X_1+\cdots+X_n)$.
16. 考虑例 4.2.3 的投资者, 假设两种股票的回报 R_1 和 R_2 的相关系数是 -1. 一个资产组合由 s_1 股第一种股票和 s_2 股第二种股票组成, 其中 s_1, $s_2\geqslant 0$. 试求资产组合使得这个资产组合的总成本是 \$6000, 回报的方差是 0. 为什么这种情况不现实?
17. 令 X 和 Y 是随机变量, 具有有限的方差. 证明: $|\rho(X,Y)|=1$ 意味着存在常数 a、b 和 c, 满足 $aX+bY=c$ 的概率是 1. 提示: 利用定理 4.6.1, 其中 $U=X-\mu_X$, $V=Y-\mu_Y$.

221

4.7 样本均值

称由独立同分布的随机变量构成的随机样本的平均值为样本均值. 样本均值是概括一个随机样本的信息的有用手段, 其方式与用均值概括一个概率分布的信息很大程度上是一样的. 在本节, 我们给出一些结论, 这些结论说明了样本均值和构成

随机样本的单个随机变量的期望值之间的联系.

4.7.1 马尔可夫不等式和切比雪夫不等式

我们以两个简单而且一般的结论作为本节的开头，分别是马尔可夫不等式和切比雪夫不等式. 然后我们将在随机样本里面应用这些不等式.

定理 4.7.1 马尔可夫不等式 假设 X 是一个满足 $\Pr(X\geqslant 0)=1$ 的随机变量，则对于每一个给定的数 $t>0$,

$$\Pr(X\geqslant t)\leqslant\frac{\mathrm{E}(X)}{t}. \tag{4.7.1}$$

证明：为简便起见，假定 X 具有离散分布，其概率函数为 f. 对于连续分布或者更加一般的分布情况下的证明类似. 对于离散分布：

$$\mathrm{E}(X)=\sum_{x}xf(x)=\sum_{x<t}xf(x)+\sum_{x\geqslant t}xf(x).$$

因为 X 只能取非负值，所以和式中所有的项都非负，因此：

$$\mathrm{E}(X)\geqslant\sum_{x\geqslant t}xf(x)\geqslant\sum_{x\geqslant t}tf(x)=t\Pr(X\geqslant t).$$ ■

马尔可夫不等式主要对于较大的 t 值有意义. 事实上，当 $t<\mathrm{E}(X)$ 时，不等式无论怎样都没什么价值，因为我们知道 $\Pr(X\leqslant t)\leqslant 1$. 然而，从马尔可夫不等式可以发现，对于每个均值为 1 的非负随机变量 X, $\Pr(X\geqslant 100)$ 的最大可能取值是 0.01. 更进一步地，能够验证当随机变量 X 满足 $\Pr(X=0)=0.99$ 和 $\Pr(X=100)=0.01$ 时，能够达到最大值.

定理 4.7.2 切比雪夫不等式 假设 X 是一个随机变量，$\mathrm{Var}(X)$ 存在，则对于每一个 $t>0$：

$$\Pr(|X-\mathrm{E}(X)|\geqslant t)\leqslant\frac{\mathrm{Var}(X)}{t^2}. \tag{4.7.2}$$

证明：令 $Y=[X-\mathrm{E}(X)]^2$，则 $\Pr(Y\geqslant 0)=1$ 并且 $\mathrm{E}(Y)=\mathrm{Var}(X)$. 对于 Y 应用马尔可夫不等式，可以得到如下结果：

$$\Pr(|X-\mathrm{E}(X)|\geqslant t)=\Pr(Y\geqslant t^2)\leqslant\frac{\mathrm{Var}(X)}{t^2}.$$ ■

由证明过程中可以看到，切比雪夫不等式只是马尔可夫不等式的一个特殊情况. 所以在证明马尔可夫不等式以后给出的那些说明也适用于切比雪夫不等式. 因为它们的一般性，这些不等式非常有用. 例如，如果 $\mathrm{Var}(X)=\sigma^2$，我们令 $t=3\sigma$，则根据切比雪夫不等式可以得到以下结果：

$$\Pr(|X-\mathrm{E}(X)|\geqslant 3\sigma)\leqslant\frac{1}{9}.$$

简单地说，该结果表明，任意给定的随机变量偏离它的均值大于或等于 3 个标准差的概率，不可能超过 1/9. 实际上，对于本书将要讨论的很多随机变量和分布来说，这个概率要比 1/9 小得多. 切比雪夫不等式的有用之处就在于它说明了对于任意 [230]

分布此概率必定小于或等于 1/9 的事实. 同时也可以证明(见本节后面的习题4)式(4.7.2)中的上界是紧的，即它不能变得更小而使得不等式对于所有分布仍然成立.

4.7.2　样本均值的性质

假设随机变量 $X_1,\cdots,X_n$ 是取自某个均值为 μ 方差为 σ^2 的分布的容量为 n 的随机样本，换句话说，假设随机变量 $X_1,\cdots,X_n$ 是独立同分布的，并且每个分布都有均值 μ 和方差 σ^2. 我们用 $\overline{X}_n$ 表示样本中这 n 个观察值的算术平均值，即：

$$\overline{X}_n=\frac{1}{n}(X_1+\cdots+X_n).$$

称该随机变量 $\overline{X}_n$ 为样本均值(sample mean).

可以很容易地计算出 $\overline{X}_n$ 的均值和方差，直接从其定义可得：

$$\mathrm{E}(\overline{X}_n)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i)=\frac{1}{n}\cdot n\mu=\mu.$$

进而，因为 $X_1,\cdots,X_n$ 是互相独立的，

$$\mathrm{Var}(\overline{X}_n)=\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}.$$

简单地说，$\overline{X}_n$ 的均值等于从中抽取随机样本的分布的均值，但 $\overline{X}_n$ 的方差只是该分布的方差的 $1/n$. 从而可以知道 $\overline{X}_n$ 的概率分布比原分布更加集中在均值 μ 周围. 换句话说，样本均值 $\overline{X}_n$ 更可能比给定分布中仅仅单个的观察值 X_i 更加接近于 μ.

如果对 $\overline{X}_n$ 运用切比雪夫不等式，此观点可以变得更加精确. 因为 $\mathrm{E}(\overline{X}_n)=\mu$ 并且 $\mathrm{Var}(\overline{X}_n)=\sigma^2/n$，从式(4.7.2)可以得到，对于每个数 $t>0$，

$$\Pr(|\overline{X}_n-\mu|\geqslant t)\leqslant\frac{\sigma^2}{nt^2}.\tag{4.7.3}$$

例4.7.1　确定必需的观测数　假设从一个均值 μ 未知但是已知标准差 σ 为2个单位的分布中抽取一个随机样本，我们来确定样本的容量为多大才能够保证 $|\overline{X}_n-\mu|$ 小于1个单位的概率至少为0.99.

231

因为 $\sigma^2=4$，由关系式(4.8.3)可以得到对于每个容量为 n 的样本：

$$\Pr(|\overline{X}_n-\mu|\geqslant 1)\leqslant\frac{4}{n}.$$

因为 n 的选取必须保证 $\Pr(|\overline{X}_n-\mu|<1)\geqslant 0.99$，从而 n 必须满足 $4/n\leqslant 0.01$，因此，得到 $n\geqslant 400$. ◀

例4.7.2　一个模拟　一个环境工程师认为供水里面含有两种污染物，砷和铅. 这两种污染物的实际浓度是独立的随机变量 X 和 Y，度量单位相同. 工程师感兴趣的是在污染物中铅的比例的平均水平，即，工程师想知道 $R=Y/(X+Y)$ 的均值. 我们假设，只要我们需要，就能够很简单地产生足够多的关于 X 和 Y 的分布的独立的伪随机(pseudo-random)数据. 一个通常的估计 $\mathrm{E}[Y/(X+Y)]$ 的方法如下：如果我们抽取 n 个样本组 $(X_1,Y_1),\cdots,(X_n,Y_n)$，并且对于 $i=1,\cdots,n$ 计算 $R_i=Y_i/(X_i+Y_i)$，则

$\overline{R}_n = \frac{1}{n}\sum_{i=1}^{n} R_i$ 是 $E(R)$ 的一个合理的估计. 为了确定 n 应该多大, 我们能像在例 4.7.1 中那样讨论, 因为已知 $|R_i| \leqslant 1$, 所以必有 $\mathrm{Var}(R_i) \leqslant 1$. (实际上, $\mathrm{Var}(R_i) \leqslant 1/4$, 但是更难证明这一点, 关于在离散情况下的证明方法, 参看本节习题 14.) 根据切比雪夫不等式, 对于每个 $\varepsilon > 0$:

$$\Pr(|\overline{R}_n - E(R)| \geqslant \varepsilon) \leqslant \frac{1}{n\varepsilon^2}.$$

因此, 如果想要 $[\overline{R}_n - E(R)] \leqslant 0.05$ 以大于或等于 0.98 的概率成立, 我们必须用 $n > 0.02/0.005^2 = 2\ 000\ 000$. ◀

需要强调的是, 在例 4.7.1 中使用切比雪夫不等式能够保证一个 $n = 400$ 的样本对于满足特定的概率要求已经足够大了, 而不用在意样本的分布的特殊类型. 如果可以得到关于此分布更进一步的信息, 则往往能够证明一个更小的 n 就足够了. 该性质将在下例中说明.

例 4.7.3 投掷硬币 假设独立地投掷一枚均匀的硬币 n 次. 对于 $i = 1, \cdots, n$, 令 $X_i = 1$ 表示第 i 次投掷得到正面, 令 $X_i = 0$ 表示第 i 次投掷得到反面. 则样本均值 $\overline{X}_n$ 就等于 n 次投掷中得到正面的次数. 我们来确定需要投掷多少次硬币才能够保证 $\Pr(0.4 \leqslant \overline{X}_n \leqslant 0.6) \geqslant 0.7$. 通过两种方法来确定这个次数: 第一种, 使用切比雪夫不等式; 第二种, 使用得到正面总次数的二项分布的精确的概率.

令 $T = \sum_{i=1}^{n} X_i$ 表示 n 次投掷中得到正面的总次数, 则 T 具有参数为 n 和 $p = 1/2$ 的二项分布. 因此, 从等式(4.2.3)可以得到 $E(T) = n/2$, 同时, 从等式(4.3.3)可以得到 $\mathrm{Var}(T) = n/4$. 因为 $\overline{X}_n = T/n$, 由切比雪夫不等式可以得到如下不等式: 232

$$\begin{aligned}\Pr(0.4 \leqslant \overline{X}_n \leqslant 0.6) &= \Pr(0.4n \leqslant T \leqslant 0.6n) \\ &= \Pr\left(\left|T - \frac{n}{2}\right| \leqslant 0.1n\right) \geqslant 1 - \frac{n}{4(0.1n)^2} = 1 - \frac{25}{n}.\end{aligned}$$

因此, 如果 $n \geqslant 84$, 此概率至少为 0.7, 满足要求.

然而, 根据本书最后给出的二项分布表, 可以发现对于 $n = 15$,

$$\Pr(0.4 \leqslant \overline{X}_n \leqslant 0.6) = \Pr(6 \leqslant T \leqslant 9) = 0.70.$$

因此, 实际上 15 次投掷就足以满足特定的概率要求. ◀

4.7.3 大数定律

例 4.3.8 中的讨论表明, 在特殊问题中, 切比雪夫不等式也许不是一个实用的用来确定样本的合适容量的工具, 因为它给出的样本容量, 可能远远大于从中抽取样本的特殊分布实际上需要的数字. 然而, 切比雪夫不等式是一个有价值的理论工具, 在这里, 将用它来证明一个重要的被称为大数定律的结论.

1. 依概率收敛

假设 $Z_1, Z_2, \cdots$ 是一个随机变量序列. 粗略地说，称这个序列收敛到给定的数 b，如果当 $n\to\infty$，Z_n 的概率分布越来越集中到 b. 正式地说，称序列 Z_1，Z_2，$\cdots$依概率收敛到 b，如果对于每个 $\varepsilon>0$：

$$\lim_{n\to\infty}\Pr(|Z_n-b|<\varepsilon)=1.$$

换句话说，如果当 $n\to\infty$ 时，Z_n 位于任意给定的围绕 b 的区间的概率趋近于1，不管该区间多么小，则该序列依概率收敛到 b.

将 $Z_1, Z_2, \cdots$依概率收敛到 b 的表述表为以下记号：

$$Z_n \xrightarrow{p} b.$$

有时候，简称为 Z_n 依概率收敛到 b.

现在，我们将要证明样本均值总是依概率收敛到从中抽取随机样本的分布的均值.

233

定理 4.7.3 大数定律 假设 $X_1,\cdots,X_n$ 是从一个均值为 μ 且方差存在的分布中抽取的随机样本，令 $\overline{X}_n$ 表示样本方差，则有：

$$\overline{X}_n \xrightarrow{p} \mu. \tag{4.7.4}$$

证明：令每个 X_i 的方差为 σ^2，根据切比雪夫不等式可得，对于每个 $\varepsilon>0$，有：

$$\Pr(|\overline{X}_n-\mu|<\varepsilon)\geqslant 1-\frac{\sigma^2}{n\varepsilon^2}.$$

因此：

$$\lim_{n\to\infty}\Pr(|\overline{X}_n-\mu|<\varepsilon)=1.$$

这意味着 $\overline{X}_n \xrightarrow{p} \mu$. ■

同样可以证明，如果从中抽取随机样本的分布具有有限的均值 μ 但是方差无限，公式(4.7.4)依然成立. 但是，关于该问题的证明超出了本书的范围.

因为 $\overline{X}_n$ 依概率收敛到 b，由此可知，如果样本的容量 n 足够大，$\overline{X}_n$ 接近于 μ 的概率很高. 因此，如果从一个均值未知的分布中抽取一个很大的样本，样本的算术平均值通常是一个对于未知的均值的比较接近的估计. 将在第5章中得到中心极限定理以后继续讨论该话题，届时，提出关于 $\overline{X}_n$ 和 μ 的差异的更精确的概率分布将成为可能.

2. 随机变量的连续函数

如果 $Z_n \xrightarrow{p} b$，并且 $g(z)$ 是一个在 $z=b$ 点处连续的函数，则很容易证明(习题15) $g(Z_n) \xrightarrow{p} g(b)$. 在我们观察有均值 μ 的随机变量，但是感兴趣的是 μ^2 或 $\log(\mu)$ 或 μ 的一些其他类型的函数的时候，这个结论有用. 类似地，容易证明如果 $Z_n \xrightarrow{p} b$，$Y_n \xrightarrow{p} c$，并且 $g(z,y)$ 在 $(z,y)=(b,c)$ 连续，则 $g(Z_n,Y_n) \xrightarrow{p} g(b,c)$(习题16).

4.7.4 扩展研究：强定律和弱定律

除了此处已经提出来的依概率收敛的概念，对于随机变量序列的收敛性还有很多其他的概念. 比如说，称随机变量序列 $Z_1, Z_2, \cdots$ *以概率 1 收敛到常数* b，如果：

$$\Pr(\lim_{n\to\infty} Z_n = b) = 1.$$

对于依概率 1 收敛的仔细研究不在本书范围之内. 能够证明如果序列 $Z_1, Z_2, \cdots$ 依概率 1 收敛到常数 b，则这个序列也依概率收敛到 b. 由于该原因，称依概率 1 收敛到常数 b 为*强收敛*(strong convergence)，而称依概率收敛为*弱收敛*(weak convergence). 为了强调这两个收敛概念的不同，这里给出的被简称为大数定律的结果经常被称为*弱大数定律*(weak law of large numbers)，而可将*强大数定律*(strong law of large numbers)描述如下：如果 $\overline{X}_n$ 是从一个均值为 μ 的分布中抽取的容量为 n 的随机样本的样本均值，则： 234

$$\Pr(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

此处不给出该结果的证明.

4.7.5 小结

大数定律说明，如果方差存在的话，随机样本的样本均值依概率收敛到单个随机变量的均值 μ. 这表明如果随机样本的容量足够大，样本均值将接近于 μ. 切比雪夫不等式提供了样本均值将以多高的概率接近于 μ 的一个粗糙的上界.

习题

1. 对于每个整数 n，令 X_n 为一个均值为 μ_n 的随机变量. 证明：如果 $\lim_{n\to\infty}\mu_n = 0$，则 $X_n \xrightarrow{p} 0$.
2. 假设 X 是一个随机变量，满足：
$$\Pr(X \geqslant 0) = 1 \text{ 并且 } \Pr(X \geqslant 10) = 1/5.$$
证明：$E(X) \geqslant 2$.
3. 假设 X 是一个随机变量，满足 $E(X) = 10$，$\Pr(X \leqslant 7) = 0.2$，并且 $\Pr(X \geqslant 13) = 0.3$.
证明：$\mathrm{Var}(X) \geqslant 9/2$.
4. 令 X 为一个随机变量，使得 $E(X) = \mu$ 并且 $\mathrm{Var}(X) = \sigma^2$，构造 X 的一个概率分布，使得：
$$\Pr(|X - \mu| \geqslant 3\sigma) = 1/9.$$
5. 从一个给定的分布中抽取的随机样本容量要多大才能保证样本均值在分布的均值的 2 倍标准差以内的概率至少为 0.99.
6. 假设 $X_1, \cdots, X_n$ 是一个从均值为 6.5，方差为 4 的分布中抽取的一个容量为 n 的随机样本. 确定 n 必须多大才能满足以下条件：
$$\Pr(6 \leqslant \overline{X}_n \leqslant 7) \geqslant 0.8.$$
7. 假设 X 是一个随机变量，满足 $E(X) = \mu$ 并且 $E(|X - \mu|^4) = \beta_4$，证明：
$$\Pr(|X - \mu| \geqslant t) \leqslant \frac{\beta_4}{t^4}.$$

8. 假设一大批产品中有30%是次品，同时假设从这批产品中抽取一个容量为 n 的随机样本. 令 Q_n 表示样本中次品的比例. 分别使用以下方法找出一个满足 $\Pr(0.2\leqslant Q_n\leqslant 0.4)\geqslant 0.75$ 的 n 的值：(a) 切比雪夫不等式；(b) 本书后面的二项分布表.
9. 令 $Z_1, Z_2, \cdots$ 为一个随机变量序列，并且假设：对于 $n=1,2,\cdots$, Z_n 的分布如下：
$$\Pr(Z_n=n^2)=\frac{1}{n}\text{并且}\Pr(Z_n=0)=1-\frac{1}{n}.$$
证明：
235
$$\lim_{n\to\infty}\mathrm{E}(Z_n)=\infty\text{但是}Z_n\xrightarrow{p}0.$$
10. 称一个随机变量序列 $Z_1, Z_2, \cdots$ 均方收敛（converge in the quadratic mean）于常数 b，如果：
$$\lim_{n\to\infty}\mathrm{E}[\,|Z_n-b|^2]=0. \tag{4.7.5}$$
证明：式(4.7.5)成立当且仅当：
$$\lim_{n\to\infty}\mathrm{E}(Z_n)=b\text{ 并且 }\lim_{n\to\infty}\mathrm{Var}(Z_n)=0.$$
提示：利用4.3节习题5.
11. 证明：如果一个随机变量序列 $Z_1, Z_2, \cdots$ 均方收敛于常数 b，则这个序列也依概率收敛到 b.
12. 设 $\overline{X}_n$ 为从一个均值为 μ、方差为 σ^2 的分布中抽取的一个容量为 n 的随机样本的样本均值，其中 $\sigma^2<\infty$. 证明：当 $n\to\infty$ 时 $\overline{X}_n$ 均方收敛于 μ.
13. 令 $Z_1, Z_2, \cdots$ 为一个随机变量序列，并且假设对于 $n=2,3,\cdots$, Z_n 的分布如下：
$$\Pr\left(Z_n=\frac{1}{n}\right)=1-\frac{1}{n^2}\text{并且}\Pr(Z_n=n)=\frac{1}{n^2},$$
a. 是否存在常数 c 使得序列依概率收敛到 c?
b. 是否存在常数 c 使得序列均方收敛到 c?
14. 设 f 为一个离散分布的分布函数，假设对于 $x\notin[0,1]$, $f(x)=0$. 证明：此分布的方差最多为4. 提示：先证明存在两点$\{0,1\}$上的一个分布，其方差至少和 f 的一样大；然后证明$\{0,1\}$上的分布的方差最多为1/4.
15. 假设 Z_n 依概率收敛到 b，并且 $g(z)$ 是一个在 $z=b$ 处连续的函数. 证明：$g(Z_n)$ 依概率收敛到 $g(b)$.
16. 假设 $Z_n\xrightarrow{p}b$, $Y_n\xrightarrow{p}c$，且 $g(z,y)$ 是一个在 $(z,y)=(b,c)$ 处连续的函数. 证明：$g(Z_n,Y_n)$ 依概率收敛到 $g(b,c)$.
17. 令 Y 为一个随机变量，其矩母函数 ψ 在0的一个开区间里有限. 证明：
$$\Pr(Y>\mu)\leqslant\min_v[\psi(v)\mathrm{e}^{-uv}].$$
236
提示：对于 $X=\mathrm{e}^{vY}$ 和 $t=\mathrm{e}^{vu}$ 应用马尔可夫不等式.

第5章 特殊分布

5.1 引 言

在本章，我们将定义和讨论几种特殊的分布，它们在概率统计中都有着广泛的应用. 这些分布主要包括一元、二元和多元的连续型和离散型分布. 一元离散型分布包括伯努利分布、二项分布、超几何分布、泊松分布、负二项分布和几何分布. 连续型一元分布包括正态分布、对数正态分布、伽玛分布、指数分布和贝塔分布，其他连续型分布有威布尔分布和帕雷托分布(将在例子和习题中介绍). 多元离散型分布主要介绍多项式分布，二元连续型分布介绍二元正态分布.

对于每一种分布，都将介绍它在实际中的应用背景及解释为何可以用它来作为某个试验的恰当的概率模型. 我们还将给出每一种分布的概率函数或概率密度函数并讨论分布的一些基本性质.

5.2 伯努利分布与二项分布

最简单的试验只有两个可能结果：0 和 1，如果以 X 记这两个可能的试验结果，那么 X 有最简单的非退化分布类型，它属于伯努利分布族. 如果 n 个相互独立的随机变量 $X_1, X_2, \cdots, X_n$ 都服从相同的伯努利分布，那么它们的和就是这些 X_i 中等于 1 的个数，而且和的分布属于二项分布族.

5.2.1 伯努利分布

一个非常简单的试验是只有两个可能结果的试验，比如正面或反面，成功或失败，有缺陷或没有缺陷，病人康复或未康复. 为方便起见，记这两个可能的结果为 0 和 1. 下面的定义就是建立在这类试验基础之上的.

如果随机变量 X 只取 0 和 1 两个值，并且相应的概率为：

$$\Pr(X=1)=p, \Pr(X=0)=1-p, \quad 0<p<1. \tag{5.2.1}$$

则称随机变量 X 服从参数为 p 的伯努利分布. 若令 $q=1-p$，则 X 的概率函数可写为：

$$f(x|p)=\begin{cases} p^x q^{1-x} & \text{对于 } x=0,1, \\ 0 & \text{其他.} \end{cases} \tag{5.2.2}$$

要证明该概率函数$f(x|p)$确实是由公式(5.2.1)所定义的伯努利分布，只要注意到$f(1|p)=p$，$f(0|p)=q$，就很容易得证.

如果X服从参数为p的伯努利分布，则由在4.3节最后所得到的结果：

$$E(X)=1\cdot p+0\cdot q=p,$$
$$E(X^2)=1^2\cdot p+0^2\cdot q=p,$$

并且，

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=pq.$$

进而，X的矩母函数为：

$$\psi(t)=E(e^{tX})=pe^t+q,\text{对于}-\infty<t<+\infty.$$

5.2.2 伯努利试验

如果无穷随机变量序列$X_1,X_2,\cdots$是独立同分布(i.i.d.)的，而且每个随机变量X_i都服从参数为p的伯努利分布，那么随机变量$X_1,X_2,\cdots$就形成参数为p的一系列伯努利试验. 同样，如果n个随机变量$X_1,\cdots,X_n$独立同分布，并且都服从参数为p的伯努利分布，则随机变量$X_1,\cdots,X_n$形成参数为p的n重伯努利试验.

下面举几个例子加以说明. 假定重复抛掷一枚均匀硬币，如果在第i次抛掷中出现正面，令$X_i=1$；如果出现反面，令$X_i=0$. 那么，随机变量$X_1,X_2,\cdots$就形成参数为$p=1/2$的一系列伯努利试验. 同样，假定由一个特定机器生产的零件中10%是有缺陷的，随机抽取n个进行观测. 如果第i个零件有缺陷，令$X_i=1$；如果没有缺陷，令$X_i=0$，$i=1,2,\cdots,n$. 那么，随机变量$X_1,\cdots,X_n$就形成参数为$p=1/10$的n重伯努利试验. 在前面几章许多医学试验的例子中(如，例4.7.5)，随机变量$X_1,X_2,\cdots$表示每个病人是否康复，在给定$P=p$的条件下，它们构成参数为p的一系列伯努利试验，其中，P是每个病人康复的未知概率.

248

5.2.3 二项分布

正如在3.1节介绍的，如果一个离散型随机变量X的概率函数为：

$$f(x|n,p)=\begin{cases}\binom{n}{x}p^xq^{n-x} & \text{当 } x=0,1,2,\cdots,n,\\ 0 & \text{其他.}\end{cases} \tag{5.2.3}$$

则称X服从参数为n和p的二项分布. 其中，n必须是正整数，$0\leqslant p\leqslant 1$. 对于各种n和p值，可以通过本书后面的附录表中查得二项分布的概率.

二项分布是概率统计中非常重要的分布，因为它具有在3.1节得到的如下结果：假定一个试验的结果是成功或者失败，独立重复进行试验n次，每次试验成功的概率为p. 如果以X表示n次试验中成功的次数，那么X服从参数为n，p的二项分布. 也可将此结果叙述如下：

如果随机变量 $X_1,\cdots,X_n$ 形成参数为 p 的 n 重伯努利试验，且 $X=X_1+\cdots+X_n$，则 X 服从参数为 n 和 p 的二项分布.

当 X 表示为 n 重伯努利试验中总的成功次数时，就很容易得到 X 的数学期望、方差和矩母函数. 在 4.2、4.3 和 4.4 节中已得到这些结果，具体如下：

$$\mathrm{E}(X)=\sum_{i=1}^{n}\mathrm{E}(X_i)=np,$$

$$\mathrm{Var}(X)=\sum_{i=1}^{n}\mathrm{Var}(X_i)=npq,$$

$$\psi(t)=\mathrm{E}(\mathrm{e}^{tX})=\prod_{i=1}^{n}\mathrm{E}(\mathrm{e}^{tX_i})=(p\mathrm{e}^t+q)^n. \tag{5.2.4}$$

读者利用(5.2.4)式的矩母函数，很容易得到 4.4 节给出的如下推论：

如果随机变量 $X_1,\cdots,X_k$ 相互独立，X_i 服从参数为 n_i 和 p 的二项分布($i=1,2,\cdots,k$)，则 $X_1+\cdots+X_k$ 服从参数为 $n=n_1+\cdots+n_k$ 和 p 的二项分布.

如果把每个 X_i 看作参数为 p 的 n_i 重伯努利试验，则很容易证明上述结果. 若 $n=n_1+\cdots+n_k$，且所有的 n 重伯努利试验都是独立的，那么和 $X_1+\cdots+X_k$ 就是参数为 p 的 n 重伯努利试验结果的和. 因此，它必定服从参数为 n 和 p 的二项分布. [249]

例 5.2.1 Castaneda v. Partida 法庭曾用二项分布计算评审团成员中来自已知的种族或宗教的概率. 在 Castaneda v. Partida 430 U. S. 482(1977)案件中，当地人口中有 79.1% 是墨西哥裔美国人. 在两年半的时间里，有 220 个人曾担任重要评审团成员，但只有 100 个人是墨西哥裔美国人. 有人宣称：在评审团重要成员的选举过程中存在着对于墨西哥裔美国人严重的歧视. 于是，法庭假定随机地、独立地选择重要评审员，并以 0.791 的概率选到墨西哥裔美国人，然后进行计算. 由于有人宣称 100 个对于墨西哥裔美国人来说太小，所以，法庭计算参数为 220 和 0.791 的二项分布随机变量 X 是 100 或更少的概率. 此概率非常小(小于 10^{-25}). 而这能够说明对墨西哥裔美国人的歧视吗？这个非常小的概率是在 X 服从参数为 220 和 0.791 的二项分布的假设下计算得到的，这意味着法庭在假定对墨西哥裔美国人没有歧视的条件下进行计算的. 换句话说，这个很小的概率是在没有歧视的前提下 $X\leqslant 100$ 的概率. 法庭更感兴趣的应该是相反方面问题的概率，即给定 $X=100$(或给定 $X\leqslant 100$)的条件下没有种族歧视的概率. 这有点像贝叶斯定理的问题，在例 5.10.1 和例 5.10.2 中，将会说明如何用贝叶斯定理计算该概率. ◀

5.2.4 小结

如果随机变量 X 的概率函数为：

$$f(x|p)=\begin{cases}p^x(1-p)^{1-x} & \text{对于 } x=0,1,\\ 0 & \text{其他.}\end{cases}$$

则 X 服从参数为 p 的伯努利分布. 如果 $X_1,\cdots,X_n$ 是相互独立的随机变量，且都服从

参数为 p 的伯努利分布，则可以将 $X_1,\cdots,X_n$ 看作一系列伯努利试验，且 $X=\sum_{i=1}^{n}X_i$ 服从参数为 n 和 p 的二项分布．可以把 X 看作是 n 重伯努利试验中成功的次数，其中在第 i 次试验中，$X_i=1$ 意味着成功，$X_i=0$ 意味着失败．

习题

1. 设随机变量 X 满足 $\mathrm{E}(X^k)=1/3, k=1,2,\cdots$．假定不可能有多于一个的分布具有类似的矩（见习题14），求 X 的分布．
2. 设随机变量 X 只能取两个值 a 和 b，且具有如下的概率：
$$\Pr(X=a)=p,\quad \Pr(X=b)=q.$$
试以与公式(5.2.2)类似的形式表达 X 的概率函数．
3. 设独立重复地抛掷一枚均匀的硬币（正面出现的概率为1/2）10次，利用本书后面附录给出的二项分布表，计算出现正面的次数大于反面次数的概率．
4. 设某一试验成功的概率为0.4，以 X 表示在15次独立试验中成功的次数．利用本书后面附录给出的二项分布表，计算概率 $\Pr(6\leqslant X\leqslant 9)$．
5. 一枚硬币，出现正面的概率为0.6，现抛掷9次．利用本书后面附录给出的二项分布表，计算出现偶数次正面的概率．

250

6. 有三个人 A，B，C，朝同一目标射击．设 A 射击3次，每次击中目标的概率为1/8，B 射击5次，每次击中目标的概率为1/4，C 射击2次，每次击中目标的概率为1/2．求目标被击中次数的期望值为多少？
7. 在习题6的条件下，假定所有的射击都是独立的．求目标被击中次数的方差为多少？
8. 某电子系统包含10个元件．设每个单独的元件失灵的概率为0.2，每个元件是否失灵是相互独立的．在至少有一个元件失灵的条件下，至少有两个元件失灵的概率为多少？
9. 假定随机变量 $X_1,\cdots,X_n$ 构成参数为 p 的 n 重伯努利试验．在给定条件
$$\sum_{i=1}^{n}X_i=k\quad(k=1,\cdots,n)$$
下，计算 $X_1=1$ 的条件概率．
10. 在一个给定家庭中，每个特定的小孩得某种遗传病的概率为 p．已知在一个有 n 个小孩的家庭中至少有一个小孩得该种遗传病，那么得该种遗传病的小孩数的期望值为多少？
11. 已知 $0\leqslant p\leqslant 1, q=1-p, n=2,3,\cdots$，计算
$$\sum_{x=2}^{n}x(x-1)\binom{n}{x}p^xq^{n-x}.$$
12. 设一个离散型随机变量 X 概率函数为 $f(x)$，则称使 $f(x)$ 达到最大值的 x 为该分布的众数．如果使 $f(x)$ 达到最大值的 x 不止一个，则称所有这样的 x 值都为该分布的众数．试求具有参数 n 和 p 的二项分布的众数．提示：考查比率 $f(x+1|n,p)/f(x|n,p)$．
13. 在一个医学试验中，病人被分成两组进行治疗，每一组5个病人．其中一组治疗成功的概率为0.5，另一组成功的概率为0.6．假定所有的病人的治疗结果是独立的．计算第一组治疗成功的人数不少于第二组的概率．

14. 在习题1中，我们假定至多一个分布具有矩：$E(X^k)=1/3, k=1,2,\cdots$. 在本题中，我们要证明确实存在这样的一个分布. 证明下面的事实并说明其隐含的意思是最多有一个分布具有上述给定的矩.
   a. $\Pr(|X|\leqslant 1)=1$.（否则，需说明 $\lim_{k\to\infty}E(X^{2k})=\infty$.）
   b. $\Pr(X^2\in\{0,1\})=1$.（否则，需证明 $E(X^4)<E(X^2)$.）
   c. $\Pr(x=-1)=0$.（否则，需证明 $E(X)<E(X^2)$.）

5.3 超几何分布

在本节中，我们将讨论不独立的伯努利随机变量. 常见的不独立伯努利随机变量产生于来自有限总体的不放回抽样. 假定一个有限总体中成功和失败的次数是已知的，如果从该总体中抽取一个样本，样本容量是固定不变的，则样本中包含的成功次数的分布属于超几何分布族.

5.3.1 超几何分布的定义

设一个箱子里有 A 个红球和 B 个蓝球，假定随机地不放回地从该箱子中取 n 个球，以 X 表示取到的红球的个数. 显然，X 既不能大于 n 也不能大于 A，因此，必定满足 $X\leqslant\min\{n,A\}$. 同样，蓝球的个数 $n-X$ 也不能大于 B，即 X 的值至少为 $n-B$. 由于 X 的值不能小于0，所以必须满足 $X\geqslant\max\{0,n-B\}$. 因此，X 的值必定是介于下列区间的整数：251

$$\max\{0,n-B\}\leqslant x\leqslant\min\{n,A\}. \tag{5.3.1}$$

以 $f(x|A,B,n)$ 表示 X 的概率函数，对于区间(5.3.1)中的每个整数 x，恰好取到 x 个红球的概率，正如例1.8.4所介绍的，可表示成：

$$f(x|A,B,n)=\frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}. \tag{5.3.2}$$

对于其他所有的 x 值，$f(x|A,B,n)=0$. 如果随机变量 X 具有这样的离散型分布，则称 X 服从参数为 A、B 和 n 的超几何分布.

例5.3.1 对观测数据集的抽样 考虑临床试验中的病人，其结果由表2-1列出. 或许需要对服用安慰剂的一组病人进行再一次检查. 假定该组包含34个病人，随机从中选择11个病人构成一个样本. 则在得到的子样本中，治疗成功(没有复发)的病人数服从什么分布呢? 以 X 表示子样本中成功的病人数. 表2-1表明在服用安慰剂的一组病人中，有10个治疗成功，24个治疗失败. 根据超几何分布的定义，X 服从参数为10、24和11的超几何分布. 特别地，X 的所有可能取值是0到10的整数. 即使子样本中只有11个病人，也不能观测到11个治疗成功的病人，因为一共只有10个成功的病人. ◀

5.3.2　超几何分布的均值和方差

再一次假定从含有 A 个红球和 B 个蓝球的箱子里随机地、不放回地抽取 n 个球. 如果第 i 次取得的是红球，令 $X_i=1$；如果第 i 次取得的是蓝球，则令 $X_i=0, i=1,\cdots,n$. 正如例4.2.4解释的，随机地抽取 n 个球，可以看作把箱子里所有的球按照一定的随机顺序进行排列，然后选取序号在前面的 n 个球. 从这个解释中可以看到，对于 $i=1,\cdots,n$，有

$$\Pr(X_i=1)=\frac{A}{A+B}, \Pr(X_i=0)=\frac{B}{A+B}.$$

因此，对于 $i=1,\cdots,n$，

252

$$E(X_i)=\frac{A}{A+B}, \mathrm{Var}(X_i)=\frac{AB}{(A+B)^2}. \tag{5.3.3}$$

因为 $X=X_1+\cdots+X_n$，则正如例4.2.4给出的，

$$E(X)=\sum_{i=1}^{n}E(X_i)=\frac{nA}{A+B}. \tag{5.3.4}$$

换句话说，具有参数 A、B 和 n 的超几何分布的均值为 $nA/(A+B)$.
进而，根据定理4.6.6，

$$\mathrm{Var}(X)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{i<j}\sum\mathrm{Cov}(X_i,X_j). \tag{5.3.5}$$

由于随机变量 $X_1,\cdots,X_n$ 的对称性，(5.3.5)式中后面一项和中间的每一项 $\mathrm{Cov}(X_i,X_j)$ 都与 $\mathrm{Cov}(X_1,X_2)$ 有相同的值. 又因为在该项和中共有 $\binom{n}{2}$ 项. 由(5.3.3)式和(5.3.5)式，可以得到

$$\mathrm{Var}(X)=\frac{nAB}{(A+B)^2}+n(n-1)\mathrm{Cov}(X_1,X_2). \tag{5.3.6}$$

如果 $n=A+B$，由于不放回地把所有的球都从盒中取出，所以，必定有 $X=A$，从而 $\mathrm{Var}(X)=0$. 于是，由(5.3.6)式，

$$\mathrm{Cov}(X_1,X_2)=-\frac{AB}{(A+B)^2(A+B-1)}.$$

因此，由(5.3.6)式，

$$\mathrm{Var}(X)=\frac{nAB}{(A+B)^2}\cdot\frac{A+B-n}{A+B-1}. \tag{5.3.7}$$

令 $T=A+B$ 表示盒子中球的总数，$p=A/T$ 表示盒子中红球所占的比例；则 $q=1-p$ 表示盒子中蓝球所占的比例. 于是，$\mathrm{Var}(X)$ 可表示成如下形式：

$$\mathrm{Var}(X)=npq\frac{T-n}{T-1}. \tag{5.3.8}$$

5.3.3 抽样方法的比较

将(5.3.8)式给出的超几何分布的方差与二项分布的方差 npq 进行比较，会发现很有趣的问题. 如果有放回地从盒子里取 n 个球，那么取到的红球个数 X 服从二项分布，且具有方差 npq. 因此，(5.3.8)式中因子 $\alpha=(T-n)/(T-1)$ 表示由于从有限总体中的不放回抽样导致的方差 $\mathrm{Var}(X)$ 减少.

如果 $n=1$，则 $\alpha=1$. 这是因为只有一个球可能被取到时，放回式抽样和不放回式抽样没有任何差别. 如果 $n=T$，像前面所提到的，$\alpha=0$，$\mathrm{Var}(X)=0$. 对于介于 1 和 T 之间的 n 值，α 的值将介于 0 和 1 之间.

253

对于每个固定的样本容量 n，可以发现当 $T\to\infty$ 时，$\alpha\to 1$. 该极限反映了当总体所含个体数 T 相对于样本容量 n 非常大时，放回式抽样和不放回式抽样几乎没有什么差别. 也就是说，样本容量 n 相对于总体 $A+B$ 微不足道时，具有参数 A，B 和 n 的超几何分布与具有参数 n 和 $p=A/(A+B)$ 的二项分布非常接近.

5.3.4 扩展研究：二项式系数定义的推广

在 1.6 节，已经给出二项式系数定义的推广，这使得超几何分布的概率函数的表达形式得以简化. 对于所有满足 $r\leqslant m$ 的正整数 r 和 m，定义二项式系数 $\binom{m}{r}$ 为：

$$\binom{m}{r}=\frac{m!}{r!\,(m-r)!}. \tag{5.3.9}$$

可以看到由(5.3.9)式定义的 $\binom{m}{r}$ 也可以写成下面的形式：

$$\binom{m}{r}=\frac{m(m-1)\cdots(m-r+1)}{r!}. \tag{5.3.10}$$

对于每个未必是正整数的实数 m 和每个正整数 r，(5.3.10)式右边的值是定义很明确的数字. 因此，对于每个实数 m 和每个正整数 r，我们可以将二项式系数 $\binom{m}{r}$ 的定义推广成(5.3.10)式的形式.

根据此定义，对于所有的正整数 r，都可以得到二项式系数 $\binom{m}{r}$ 的值. 如果 $r\leqslant m$，则 $\binom{m}{r}$ 的值由公式(5.3.9)给出；如果 $r>m$，则 $\binom{m}{r}=0$. 对于所有的实数 m，我们定义 $\binom{m}{0}=1$.

根据二项式系数的推广定义，可以看到，对于满足 $x>A$ 或者 $n-x>B$ 的每个整

数 x，都有$\binom{A}{x}\binom{B}{n-x}=0$. 因此，可以将具有参数 A、B 和 n 的超几何分布的概率函数写成如下形式：

$$f(x|A,B,n)=\begin{cases}\dfrac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} & \text{对于 } x=0,1,\cdots,n,\\ 0 & \text{其他.}\end{cases} \tag{5.3.11}$$

254 根据(5.3.10)式，$f(x|A,B,n)>0$ 当且仅当 x 是位于区间(5.3.1)上的整数时.

5.3.5 小结

本节，我们介绍了超几何分布族. 假定一个有限总体含有 T 个个体，其中代表成功的个体数为 A，失败的个体数则为 $B=T-A$. 从该总体中不放回地抽取 n 个个体，以 X 表示样本中代表成功的个体数，则 X 的分布就是具有参数 A、B 和 n 的超几何分布. 同时，又对二项式系数进行了推广，如此一来，对于所有的实数 m 和正整数 r，$\binom{m}{r}$都有定义.

习题

1. 在例 5.3.1 中，来自于服用安慰剂的一组的 11 个病人构成一个子样本，计算所有治疗成功的 10 个病人都出现在该子样本中的概率.
2. 假设一个盒子中有 5 个红球和 10 个蓝球. 现从该盒中不放回地随机抽取 7 个球. 求至少取到 3 个红球被取到的概率.
3. 假设一个盒子中有 5 个红球，10 个蓝球. 现从该盒中不放回地随机抽取 7 个球. 以 $\overline{X}$ 表示样本中红球所占的比例，$\overline{X}$ 的均值和方差分别为多少?
4. 如果随机变量 X 服从参数为 $A=8$、$B=20$ 和 n 的超几何分布，使 $\text{Var}(X)$ 达到最大值的 n 为多少?
5. 设一个班级有 T 名学生，其中 A 名是男生，$T-A$ 名是女生，现从该班不重复地随机挑选 n 名学生，以 X 表示挑到的男生数. 求使 $\text{Var}(X)$ 达到最大值的样本容量 n 为多少?
6. 设随机变量 X_1 和 X_2 相互独立，且 X_1 服从参数为 n_1 和 p 的二项分布，X_2 服从参数为 n_2 和 p 的二项分布，其中的 p 是相同的. 对于任意固定的 $k(k=1,2,\cdots,n_1+n_2)$，在给定 $X_1+X_2=k$ 的条件下，求 X_1 的条件分布.
7. 假定在一大堆包含 T 个生产零件中，30% 是有缺陷的，70% 没有缺陷. 从中无重复地随机抽取 10 个零件. 写出 (a) 不多于一个有缺陷的零件被取到的概率的确切表达式；(b) 利用二项分布写出该概率的近似表达式.
8. 设一个总体中含有 T 个人，a_1，$\cdots$，a_T 表示这 T 个人的身高. 假定从中无重复地随机选取 n 个人，以 X 表示这 n 个人身高的和. 计算 X 的均值和方差.

9. 求$\binom{3/2}{4}$的值.

10. 证明：对于所有的正整数 n，k，有

$$\binom{-n}{k}=(-1)^k\binom{n+k-1}{k}.$$

5.4 泊松分布

许多试验都需要观测随机事件出现的次数. 该种例子有：服务的顾客数，某交换台收到的电话呼叫次数，洪水及其他自然灾害和人为因素导致的事故发生的次数等等. 泊松分布族用来描述在固定时间间隔里来访者的数目，它也可用来作为具有非常小的成功概率的二项分布的近似.

255

5.4.1 泊松分布的定义和性质

1. 概率函数

设 X 是一离散型随机变量，且 X 的取值为所有非负整数. 如果 X 的概率函数具有如下形式：

$$f(x|\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!} & \text{对于 } x=0,1,2,\cdots,\\ 0 & \text{其他.}\end{cases}\tag{5.4.1}$$

则称 X 服从均值为 $\lambda(\lambda>0)$ 的泊松分布.

显然，对于所有的 x 值，$f(x|\lambda)\geqslant 0$. 要证明由(5.4.1)式定义的 $f(x|\lambda)$ 满足概率函数的定义，必须证明 $\sum_{x=0}^{\infty}f(x|\lambda)=1$. 由初等计算可知，对于任意实数 λ，有

$$e^{\lambda}=\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}.\tag{5.4.2}$$

于是，

$$\sum_{x=0}^{\infty}f(x|\lambda)=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{-\lambda}e^{\lambda}=1.\tag{5.4.3}$$

2. 均值和方差

上面已经指出，称具有(5.4.1)式给出的概率函数的分布为具有均值 λ 的泊松分布. 为了证明该定义，必须说明 λ 就是此分布的均值. 均值 $E(X)$ 由下述无穷级数所确定：

$$E(X)=\sum_{x=0}^{\infty}xf(x|\lambda)$$

由于 $x=0$ 时，级数的第一项 $xf(x|\lambda)=0$，所以可以将该项删掉，因此，级数的和从

$x=1$ 项开始. 因此,

$$E(X)=\sum_{x=1}^{\infty} xf(x|\lambda)=\sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}=\lambda\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}.$$

若在和号中令 $y=x-1$，则

$$E(X)=\lambda\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!}.$$

根据(5.4.3)式，上述级数的和等于 1，因此，$E(X)=\lambda$.

可以用与上面同样的技巧得到泊松分布的方差. 首先，考察下述期望:

$$E[X(X-1)]=\sum_{x=0}^{\infty}x(x-1)f(x|\lambda)=\sum_{x=2}^{\infty}x(x-1)f(x|\lambda)$$

256

$$=\sum_{x=2}^{\infty}x(x-1)\frac{e^{-\lambda}\lambda^x}{x!}=\lambda^2\sum_{x=2}^{\infty}\frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!}.$$

若令 $y=x-2$，则

$$E[X(X-1)]=\lambda^2\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!}=\lambda^2. \tag{5.4.4}$$

由于 $E[X(X-1)]=E(X^2)-E(X)=E(X^2)-\lambda$，由(5.4.4)式，得 $E(X^2)=\lambda^2+\lambda$. 因此,

$$\mathrm{Var}(X)=E(X^2)-[E(X)]^2=\lambda. \tag{5.4.5}$$

如此一来，概率函数为(5.4.1)式的泊松分布，均值和方差都为 λ.

3. 矩母函数

下面，我们将给出由概率函数为(5.4.1)式的泊松分布的矩母函数 $\psi(t)$. 对于每个 t 值($-\infty<t<\infty$)，

$$\psi(t)=E(e^{tX})=\sum_{x=0}^{\infty}\frac{e^{tx}e^{-\lambda}\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}.$$

由(5.4.2)式，对于 $-\infty<t<\infty$，

$$\psi(t)=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}. \tag{5.4.6}$$

由(5.4.6)式给出的矩母函数可以求得均值、方差和其他矩. 此处，我们不推导其他的矩，但利用矩母函数得到泊松分布的下列性质.

定理 5.4.1　如果随机变量 $X_1,\cdots,X_k$ 相互独立，而且 X_i 服从均值为 $\lambda_i(i=1,\cdots,k)$ 的泊松分布，则 $X_1+\cdots+X_k$ 服从均值为 $\lambda_1+\cdots+\lambda_k$ 的泊松分布.

证明: 以 $\psi_i(t)$ 表示 $X_i(i=1,\cdots,k)$ 的矩母函数，$\psi(t)$ 表示 $X_1+\cdots+X_k$ 的矩母函数. 由于 $X_1,\cdots,X_k$ 相互独立，于是，对于 $-\infty<t<\infty$，有

$$\psi(t)=\prod_{i=1}^{k}\psi_i(t)=\prod_{i=1}^{k}e^{\lambda_i(e^t-1)}=e^{(\lambda_i+\cdots+\lambda_k)(e^t-1)}.$$

由(5.4.6)式可以看出，矩母函数 $\psi(t)$ 就是均值为 $\lambda_1+\cdots+\lambda_k$ 的泊松分布的矩母函数. 因此，$X_1+\cdots+X_k$ 服从均值为 $\lambda_1+\cdots+\lambda_k$ 的泊松分布. 定理得证. ■

在本书后面的附录表中给出具有不同均值 λ 的泊松分布的概率表.

例 5.4.1 顾客到达 一个店主认为平均每小时到本店购物的顾客数为 4.5. 店主用独立地并且服从均值为 4.5 的泊松随机变量来刻划在互不相交的不同时间区间里来访的顾客数. 在两小时的时间间隔里至少有 12 名顾客到本店购物的概率为多少? 以 X_1 表示第一个小时内来访的顾客数, X_2 表示第二个小时内来访的顾客数. 店主认为 X_1 和 X_2 是独立的泊松随机变量, 且均值都是 4.5. 根据定理 5.4.1, 两个小时内来访的顾客总数 $X = X_1 + X_2$ 服从均值为 9 的泊松分布. 利用本书后面给出的泊松概率表, 将 $\lambda = 9$ 一列对应的数值加起来. 或者将 $k = 0, \cdots, 11$ 对应的数值加起来, 然后用 1 减去它们的和或者从 $k = 12$ 加到最后. 无论哪一种方法, 得到的结果都是 $\Pr(X \geqslant 12) = 0.1970$. ◀

257

5.4.2 二项分布的泊松近似

现在来证明当 n 很大且 p 接近于 0 时, 可以用均值为 np 的泊松分布来近似具有参数 n 和 p 的二项分布. 假定随机变量 X 服从参数为 n 和 p 的二项分布, 对每个 x 值令 $\Pr(X = x) = f(x|n,p)$, 由(5.2.3)式, 对于 $x = 1, 2, \cdots, n$, 有

$$f(x|n,p) = \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x}.$$

若令 $\lambda = np$, 则可将 $f(x|n, p)$ 重新写成下述形式:

$$f(x|n,p) = \frac{\lambda^x}{x!}\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-x+1}{n}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}. \tag{5.4.7}$$

令 $n \to \infty$, $p \to 0$, 且使 np 在整个极限过程中保持固定的值 λ. 由于随着 $n \to \infty$, λ 和 x 是固定不变的, 则

$$\lim_{n\to\infty} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-x+1}{n}\left(1-\frac{\lambda}{n}\right)^{-x} = 1.$$

而且, 由初等计算知,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \mathrm{e}^{-\lambda}. \tag{5.4.8}$$

由(5.4.7)式, 对于每个固定的正整数 x,

$$f(x|n,p) \to \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}. \tag{5.4.9}$$

最后, 当 $x = 0$ 时,

$$f(x|n,p) = (1-p)^n = \left(1-\frac{\lambda}{n}\right)^n.$$

于是, 根据(5.4.8)式, (5.4.9)式对于 $x = 0$ 也成立. 因此, 对于任意的非负整数 x, (5.4.9)式都成立.

在(5.4.9)式的右边, $f(x|\lambda)$ 是均值为 λ 的泊松分布的概率. 因此当 n 很大, p 接近于 0 时, 对于 $x = 0, 1, \cdots$, 二项分布的概率值 $f(x|n,p)$ 可以由参数为 $\lambda = np$ 的泊

258

松分布的概率值$f(x|\lambda)$来近似.

例 5.4.2　概率的近似计算　假设在一个大的总体中，患有某种疾病的人群所占比例为0.01. 现在来计算随机抽取的200个人中，至少有4个人患有该疾病的概率.

在此例中，可以假定随机抽取的200个人中，患有该疾病的人数的精确分布为二项分布，其参数为$n=200$和$p=0.01$. 因此，可以用参数为$\lambda=np=2$的泊松分布来近似. 如果以X表示服从泊松分布的随机变量，那么根据本书后面的附录表，可得$\Pr(X\geqslant 4)=0.1428$. 于是，至少有4个人患有该疾病的概率近似值为0.1428，而精确值为0.1420. ◀

5.4.3　泊松过程

在例5.4.1中，店主认为任意一小时内来访的顾客数服从均值为4.5的泊松分布. 如果店主感兴趣的是半小时或4小时15分钟内来访的顾客数，那么均值为多少呢？是不是可以认为半小时内来访的顾客数的均值为2.25？如果店主作出附加假设：顾客是按照平均一小时4.5个的泊松过程光顾本店的，则答案是肯定的. 单位时间里速率为λ的泊松过程，是满足下列两条性质的过程：

i. 在固定的时间间隔t内，来访者的数量服从参数为λt的泊松分布；

ii. 在两个不相交的时间区间内，来访者的数量是相互独立的.

虽然前面已经介绍过泊松过程是用来描述在一定时间间隔内来访者的数量，实际上，泊松过程还有更一般的描述. 例如，不但可以用泊松分布描述一定时间间隔内发生事件的次数，也可以描述一定空间内发生事件的次数. 可以用泊松过程来描述电话交换台收到的电话呼叫次数，某一放射源放射的原子颗粒数，森林里受灾的树木数，一件合成产品表面的瑕疵数等等. 泊松过程的模型之所以有如此广泛的用途，主要有两方面的原因：一是，模型计算起来比较简便；二是，如果关于事件的发生可以作出三条似乎合理的假设，那么对此模型就有一个很好的数学证明. 在几个例子之后，将详细介绍这三条假设.

例 5.4.3　放射性颗粒　假定一放射性颗粒，按照泊松过程，以每分钟3个颗粒的平均速率射中某一目标. 现在来确定在两分钟内至少有10个颗粒击中目标的概率.

在泊松过程中，在任何一分钟的时间间隔内，击中目标的粒子数服从均值为λ的泊松分布. 由于在一分钟的时间间隔内，平均击中目标的颗粒数为3，所以，在此例中，服从$\lambda=3$的泊松分布，在两分钟的时间间隔内击中目标的颗粒数服从均值为6的泊松分布. 由本书后面的泊松分布表可以得到$\Pr(X\geqslant 10)=0.0838$. ◀

259

例 5.4.4　饮用水中的似隐孢菌素　似隐孢菌素是存在于很小的卵囊中的一种原生动物，如果吸食了它，将引起疼痛甚至死亡. 公共供水系统中偶尔会检测到卵囊. 每五升水中存在一个卵囊的低浓度就足够引起沸水发出警戒信号. 1993年4月威斯康星州米尔沃基市，在发生芽孢生殖时，成千上万的人生病. 不同的供水系统有不

同的控制喝水过程中出现原生动物的方法. 实际上，控制系统运用的是检测技术，但其中存在的一个问题是，该项技术并不是很灵敏. 一项普遍应用的技术是，让大量的水通过一个非常精细的过滤器，然后通过某种方法处理在过滤器上的残留物质，以确认似隐孢菌素卵囊数. 同时，计算和记录下卵囊的数目. 即使在过滤器上有一个卵囊，也只有 0.1 的概率可以记录下来.

假定在一个特定的供水系统中，卵囊是按照每升水 λ 个卵囊的泊松过程出现的. 设过滤系统能够过滤出所有的卵囊，但事实上，计数系统却只能以概率 p 观测到留在过滤器上的卵囊. 假定计数系统观测或者没有观测到每个卵囊是相互独立的. 那么，t 升水中，观测到的卵囊数服从怎样的分布呢?

以 Y 表示 t 升水中含有的卵囊数(假定所有的卵囊都残留在过滤器表面)，则 Y 服从参数为 λt 的泊松分布. 若第 i 个卵囊被观测到，令 $X_i=1$，否则，令 $X_i=0$. 以 X 表示观测到的卵囊数，如果 $Y=y$，则 $X=X_1+\cdots+X_y$. 在 $Y=y$ 的条件下，X_i 是具有参数为 p 并且相互独立的伯努利随机变量. 因此，在 $Y=y$ 的条件下，X 服从参数为 y 和 p 的二项分布. 我们希望得到 X 的边际分布，这可以通过随机变量的全概率公式(3.6.11)得到. 对于 $x=0,1,\cdots$,

$$
\begin{aligned}
f_1(x) &= \sum_{y=0}^{\infty} g_1(x|y)f_2(y) \\
&= \sum_{y=x}^{\infty} \binom{y}{x} p^x(1-p)^{y-x}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^y}{y!} \\
&= \mathrm{e}^{-\lambda t}\frac{(p\lambda t)^x}{x!}\sum_{y=x}^{\infty}\frac{[\lambda t(1-p)]^{y-x}}{(y-x)!} \\
&= \mathrm{e}^{-\lambda t}\frac{(p\lambda t)^x}{x!}\sum_{u=0}^{\infty}\frac{[\lambda t(1-p)]^{u}}{u!} \\
&= \mathrm{e}^{-\lambda t}\frac{(p\lambda t)^x}{x!}\mathrm{e}^{\lambda t(1-p)} = \mathrm{e}^{-p\lambda t}\frac{(p\lambda t)^x}{x!}.
\end{aligned}
$$

很容易看出，上述式子是均值为 $p\lambda t$ 的泊松分布的概率分布. 卵囊计数丢失 $(1-p)$ 比例的影响低于在泊松过程中将参数 λ 变为 $p\lambda$ 的影响.

设 $\lambda=0.2$，$p=0.1$，则要保证至少以 0.9 的概率观测到至少一个卵囊，必须过滤多少升的水? 至少观测到一个卵囊的概率为 1 减去没有观测到任何卵囊的概率 $\mathrm{e}^{-p\lambda t}=\mathrm{e}^{-0.02t}$. 于是，需要足够大的 t 使得 $1-\mathrm{e}^{-0.02t}\geq 0.9$，即 $t\geq 115$. 一个典型的过程就是检测 100 升水，有 $1-\mathrm{e}^{-0.02\times 100}=0.86$ 的概率检测到至少一个卵囊. ◀

260

5.4.4 扩展研究：泊松过程模型的假设

接下来，我们将考虑时间区间，但所作的假设也可以同样很好地用在二维或者三维的子区域，或者直线上的子区间. 事实上，可以用泊松过程来描述任何区域上事件发生的模型，只要这些区域可以分解成任意小的子区域.

第一个假设是在互不相交的时间区间上，事件的发生是相互独立的. 比如，在特定时间里，即使电话交换台收到的电话呼叫次数非常多，但在接下来的时间间隔内至少有一次呼叫的概率不受任何影响. 同样，即使在一个很长的时间间隔内电话交换台上没有任何呼叫，在紧接着的很短的时间间隔内有一次呼叫的概率仍然不受影响.

第二个假设是，在一个非常短的时间间隔内发生一个事件的概率必须与时间长度成正比. 为了将这个条件更形象地表达出来，我们利用标准的数学符号，以 $o(t)$ 表示 t 的函数，并且具有下列性质：

$$\lim_{t \to 0} \frac{o(t)}{t} = 0. \tag{5.4.10}$$

根据(5.4.10)式，当 $t \to 0$ 时，函数 $o(t)$ 必须趋近于0，而且，趋近于0的速度必须比 t 还要快. 比如，函数 $o(t) = t^{\alpha}$，$\alpha > 1$ 就满足该条件. 可以证明该函数满足(5.4.10)式. 也可将第二个条件表示成：存在一个常数 $\lambda > 0$，使得在每个长度为 t 的时间间隔内，至少有一个事件发生的概率具有形式 $\lambda t + o(t)$. 如此一来，对于每个非常小的时间 t，在每个长度为 t 的时间间隔内，至少有一个事件发生的概率等于 λt 加上一个无穷小量.

第二个假设的一个结果是在整个观测期间，观测过程是平稳的，即，整个观测期间事件发生的概率是相同的. 既没有事件发生频率较高的区间，也没有事件发生频率较低的区间. 该条件反映了这样一个事实：同一个常数 λ 表示了在整个观测期间的每个时间间隔内事件发生的概率. 可以通过更复杂的数学进一步放宽第二个假设，但这里不做介绍.

第三个假设是，对于每个非常短的时间间隔，至少有两个事件发生的概率具有比恰有一个事件发生的概率更高阶的无穷小量. 用符号表示，在长度为 t 的时间间隔内，至少有两个事件发生的概率为 $o(t)$. 因此，在一个非常短的时间间隔内，至少有两个事件发生的概率相对于发生一个事件的概率可以忽略不计. 当然，根据第二个假设，在相同的时间间隔内，发生一个事件的概率相对于没有任何事件发生的概率也可以忽略不计.

261

可以证明，在上述三条假设下的过程满足速率为 λ 的泊松过程定义. 具体证明方法见本节习题16.

5.4.5 小结

常用泊松分布来描述事件出现的次数，用速率为 λ 的泊松过程来描述单位时间内（或单位区域内）期望出现的事件数为 λ 的随机事件出现的次数. 假定在互不相交的时间间隔内（或互不相交的区域内）事件的发生是相互独立的，且两个甚至更多事件不能在同一时间（或同一地点）发生，则时间间隔（或区域大小）为 t 时，事件出现的次数服从均值为 λt 的泊松分布. 如果 n 较大，p 较小，则可以用均值为 np 的泊松分布来近似参数为 n 和 p 的二项分布.

习题

1. 在例5.4.4中，设 $\lambda=0.2$，$p=0.1$，计算过滤100升水后，至少检测到两个卵囊的概率.
2. 假定在某个周末，在某个特定的十字路口，发生交通事故的次数服从均值为0.7的泊松分布. 则周末在该十字路口至少发生3次交通事故的概率为多少？
3. 设某个特定的生产过程生产的一卷布中有缺陷的数量服从均值为0.4的泊松分布. 现随机抽取5卷布进行观测，则5卷布中有缺陷的总数至少是6的概率为多少？
4. 假定在某本书中，平均每页有 λ 处印刷错误，且错误的出现服从泊松过程. 求在特定的某页上没有印刷错误的概率为多少？
5. 设一本书有 n 页，平均每页有 λ 处印刷错误. 则至少有 m 页印刷错误多于 k 处的概率为多少？
6. 设某种磁带每1000英尺中平均有3处缺陷. 则一卷1200英尺的磁带中没有缺陷的概率为多少？
7. 设某个商店平均每小时服务15名顾客，则在两小时的时间间隔内至少服务20名顾客的概率为多少？
8. 设随机变量 X_1 和 X_2 相互独立，且 X_i 服从参数为 $\lambda_i(i=1,2)$ 的泊松分布. 对于固定的 $k(k=1,2,\cdots)$，在给定 $X_1+X_2=k$ 的条件下，求 X_1 的条件分布.
9. 设某台机器生产的零件总数服从均值为 λ 的泊松分布，所有的零件相互独立，且每个零件有缺陷的概率为 p. 试求由此台机器生产的零件中有缺陷的零件数的边际概率分布. [262]
10. 在习题9中，以 X 表示有缺陷的零件数，Y 表示没有缺陷的零件数. 证明：X 和 Y 是相互独立的.
11. 利用5.2节习题12众数的定义，求具有均值为 λ 的泊松分布的众数.
12. 假定在某个特定的人群中，色盲所占比例为0.005. 从该人群中，随机地选取600人，则选到的色盲数不多于一个的概率为多少？
13. 在婴儿出生中，出现三胞胎的概率近似为0.001. 则在一所大型医院里，700个新生婴儿中恰有一个三胞胎的概率为多少？
14. 某航班出售200张机票，但飞机上只有198个座位. 这是因为在机票购买者中，平均有1%的人不会乘坐该次航班. 求乘坐该次航班的所有人均有座位的概率.
15. 设互联网使用者按速率 λ/小时登录某特定网站，但 λ 是未知的. 网站维护者认为 λ 具有连续型分布，其概率密度函数为：

$$f(\lambda)=\begin{cases}2e^{-2\lambda} & \text{对于 } \lambda>0,\\ 0 & \text{其他.}\end{cases}$$

   设 X 表示一小时内登录该网站的人数. 如果观测到 $X=1$，试求在 $X=1$ 的条件下 λ 的概率密度函数.
16. 在此题中，将证明在泊松过程的三条假设下事件的发生确实是按照泊松过程进行的. 只要证明对于任意的 t，在长度为 t 的时间间隔内，事件发生的次数服从均值为 λt 的泊松分布. 设 X 表示在长度为 t 的时间间隔内事件发生的次数. 证明过程中，可以利用公式(5.4.8)得以下：

$$\lim_{u\to 0}(1+au+o(u))^{1/u}=\mathrm{e}^{a},\quad a\text{为任意的实数.}$$

a. 对于每个正整数 n，将时间间隔 t 分成 n 个互不相交的长度为 t/n 的时间间隔. 如果在第 i 个时间间隔内恰有一个事件发生，令 $Y_i=1$，$i=1,\cdots,n$. 以 A_i 表示第 i 个时间间隔内出现两次及两次以上的事件，令 $W_n=\sum_{i=1}^{n}Y_i$. 对于每个非负整数 k，证明：$\Pr(X=k)=\Pr(W_n=k)+\Pr(B)$，其中，$B$ 是 $\bigcup_{i=1}^{n}A_i$ 的子集.

b. 证明：$\lim_{n\to\infty}\Pr(\bigcup_{i=1}^{n}A_i)=0$. 提示：只须证明 $\Pr(\bigcap_{i=1}^{n}A_i^{c})=(1+o(u))^{1/u}$，其中 $u=1/n$.

c. 证明：$\lim_{n\to\infty}\Pr(W_n=k)=\mathrm{e}^{-\lambda}(\lambda t)^{k}/k!$. 提示：$\lim_{n\to\infty}n!/[n^{k}(n-k)!]=1$.

d. 证明：X 服从均值为 λt 的泊松分布.

5.5 负二项分布

由前面可知，在成功的概率为 p 的 n 重伯努利试验中，成功的次数服从参数 n 和 p 的二项分布. 经常会考虑这样的试验：试验进行到出现一定次数的成功为止，而不是在固定试验次数中考虑成功出现的次数. 比如说，有一台设备需要维修，可以让它一直运转，直到出现一定次数的故障为止再进行维修. 则直到出现一定次数的成功为止失败总次数的分布属于负二项分布族.

5.5.1 负二项分布的定义

在一个独立的无穷试验序列中，每次试验只有成功或失败两种结果，并设在每次试验中成功的概率为 $p(0<p<1)$，则失败的概率为 $q=1-p$. 这样的一系列试验构成参数为 p 的伯努利试验序列. 在本节，我们将讨论在恰好出现 r 次成功前失败总次数的分布，其中 r 是固定的正整数.

263

对于 $n=r,r+1,\cdots$，令 A_n 表示 r 次成功前需试验的总次数为 n 这一事件. 正如在例2.2.7解释的，事件 A_n 发生当且仅当前 $n-1$ 次试验中恰有 $r-1$ 次成功出现且第 r 次成功发生在第 n 次试验. 由于所有的试验是相互独立的，因此，有

$$\Pr(A_n)=\binom{n-1}{r-1}p^{r-1}q^{(n-1)-(r-1)}\cdot p=\binom{n-1}{r-1}p^{r}q^{n-r}.\tag{5.5.1}$$

对于每个 x 值($x=0,1,2,\cdots$)，在出现第 r 次成功之前恰有 x 次失败出现的事件相当于成功 r 次前进行的试验总次数为 $r+x$ 的事件. 即，如果以 X 表示第 r 次成功前失败的次数，则 $\Pr(X=x)=\Pr(A_{r+x})$. 如果将 $\Pr(X=x)$ 记为 $f(x|r,p)$，则由(5.5.1)式，得

$$f(x|r,p)=\begin{cases}\dbinom{r+x-1}{x}p^{r}q^{x} & \text{对于 } x=0,1,2,\cdots,\\ 0 & \text{其他.}\end{cases}\tag{5.5.2}$$

如果离散型随机变量 X 具有(5.5.2)式所表示的概率函数 $f(x|r,p)$，则称 X 服从参数为 r 和 $p(r=1,2,\cdots$ 且 $0<p<1)$ 的负二项分布.

5.5.2 几何分布

负二项分布中，当 $r=1$ 时，称作几何分布. 即，如果离散型随机变量 X 的概率函数 $f(x|1,p)$ 具有形式：

$$f(x|1,p)=\begin{cases}pq^x & \text{对于 } x=0,1,2,\cdots,\\ 0 & \text{其他.}\end{cases} \tag{5.5.3}$$

则称随机变量 X 服从参数为 $p(0<p<1)$ 的几何分布.

考虑无穷伯努利试验序列，每次试验结果为成功或失败，且每次成功的概率为 p. 如果以 X_1 表示在第一次成功前失败的次数，则 X_1 服从参数为 p 的几何分布.

更一般地，对于 $j=2,3,\cdots$，令 X_j 表示出现 $j-1$ 次成功后至第 j 次成功之间失败的次数. 由于所有的试验是相互独立的，每次成功的概率均为 p，则 X_j 服从参数为 p 的几何分布，且 $X_1,X_2,\cdots$ 相互独立. 对于 $r=1,2,\cdots$，$X_1+\cdots+X_r$ 表示恰好出现 r 次成功前失败的总次数. 因此，$X_1+\cdots+X_r$ 服从参数为 r 和 p 的负二项分布. 于是，我们得到如下结论： 264

若随机变量 $X_1,\cdots,X_r$ 独立同分布，且每个 X_i 均服从参数为 p 的几何分布，则 $X_1+\cdots+X_r$ 服从参数为 r 和 p 的负二项分布.

5.5.3 负二项分布和几何分布的其他性质

1. 矩母函数

如果 X_1 服从参数为 p 的几何分布，则 X_1 的矩母函数 $\psi(t)$ 为：

$$\psi_1(t)=\mathrm{E}(\mathrm{e}^{tX_1})=p\sum_{x=0}^{\infty}(q\mathrm{e}^t)^x. \tag{5.5.4}$$

对于满足 $0<q\mathrm{e}^t<1$ 的每个 t，即 $t<\log(1/q)$，(5.5.4)式中的无穷级数都为有限值. 由初等计算可知，对于每个 $\alpha(0<\alpha<1)$，有

$$\sum_{x=0}^{\infty}\alpha^x=\frac{1}{1-\alpha}.$$

因此，对于 $t<\log(1/q)$，

$$\psi_1(t)=\frac{p}{1-q\mathrm{e}^t}. \tag{5.5.5}$$

由定理4.4.3可知，如果随机变量 $X_1,\cdots,X_r$ 独立同分布，且每个随机变量的矩母函数均为 $\psi_1(t)$，则 $X_1+\cdots+X_r$ 的矩母函数为 $[\psi_1(t)]^r$. 由于 $X_1+\cdots+X_r$ 服从参数为 r 和 p 的负二项分布，则可以得到如下结果：

如果随机变量 X 服从参数为 r 和 p 的负二项分布，则 X 的矩母函数为：

$$\psi(t)=\left(\frac{p}{1-q\mathrm{e}^t}\right)^r \quad \text{对于 } t<\log\left(\frac{1}{q}\right). \tag{5.5.6}$$

2. 均值和方差

如果随机变量 X_1 服从参数为 p 的几何分布，则可以通过对(5.5.4)式的矩母函数求导来得到 X_1 的均值和方差，其结果为：

$$E(X_1)=\psi'_1(0)=\frac{q}{p} \tag{5.5.7}$$

265

$$\mathrm{Var}(X_1)=\psi''_1(0)-[\psi'_1(0)]^2=\frac{q}{p^2}. \tag{5.5.8}$$

现在假定 X 服从参数为 r 和 p 的负二项分布，如果 X 可以表示成 r 个独立随机变量的和 $X_1+\cdots+X_r$，且每个随机变量均与 X_1 具有相同的分布，则由(5.5.7)式和(5.5.8)式，X 的均值和方差必定是：

$$E(X)=\frac{rq}{p},\mathrm{Var}(X)=\frac{rq}{p^2}. \tag{5.5.9}$$

例 5.5.1 彩票中的三重数 一个常见的彩票博弈中，涉及每天独立地、可重复地从0到9取3个数字. 当所有的数字相同时，该事件称作三重数，彩票观察者看到这一事件时，经常会很兴奋，如果以 p 表示观察到三重数的概率，X 表示出现第一个三重数前没有三重数出现的天数，则 X 服从几何分布. 在本例中，很容易看到 $p=0.01$，这是因为每天都相当于从1000种可能结果中等可能地抽取10种不同的三重数. 因此，直到观察到三重数的期望天数为 $1+E(X)=100$. ◀

3. 几何分布的无记忆性

我们继续考虑无穷伯努利的试验序列，每次试验结果为成功或失败，且每次成功的概率为 p，则出现第一次成功前失败的次数服从几何分布. 现假定前20次试验出现的结果均为失败，由于所有的试验是相互独立的，则在出现第一次成功前额外的失败次数服从参数为 p 的几何分布. 事实上，该过程相当于从第21次试验重新开始，前20次试验中出现的一系列失败对于以后的试验没有任何影响. 通常，称此性质为几何分布的无记忆性.

在试验的开始，出现第一次成功前失败的期望次数为 q/p，正如(5.5.7)式给出的. 如果前20次试验均为失败，则出现第一次成功前失败的期望总次数为 $20+(q/p)$. 用数学式子表示几何分布的无记忆性为：

如果 X 服从参数为 p 的几何分布，则对于所有的非负整数 k、t，有

$$\Pr(X=k+t|X\geqslant k)=\Pr(X=t). \tag{5.5.10}$$

可以利用(5.5.3)式给出的概率分布 $f(x|1,p)$ 得到(5.5.10)式简单的数学证明. 在本节的习题8将要求给出详细证明.

5.5.4 扩展研究：负二项分布定义的推广

对于每个 $r>0$（未必是整数）和每个满足 $0<p<1$ 的 p，运用(5.3.10)式给出

266

的二项式系数的定义，可以将函数 $f(x|r,p)$ 看作是离散型随机变量的概率分布. 即，

对于 $r>0$，$0<p<1$，

$$\sum_{x=0}^{\infty}\binom{r+x-1}{x}p^r q^x=1. \tag{5.5.11}$$

5.5.5 小结

对具有成功概率为 p 的独立伯努利试验序列进行观测，则直到出现第 r 次成功前所失败的次数服从参数为 r 和 p 的负二项分布. 特例是 $r=1$ 时，称为参数为 p 的几何分布. 具有相同参数 p 且相互独立的负二项分布的和仍为负二项分布.

习题

1. 考虑例 5.5.1 中描述的彩票发行的例子.
   a. 计算在特定的两天里都出现三重数的概率.
   b. 设在某一天里观测到三重数，计算在接下来的一天仍然观测到三重数的概率.
2. 在独立抛掷硬币的序列中，假设每次抛掷出现正面的概率为 1/30.
   a. 出现 5 次正面前出现反面的期望次数为多少？
   b. 出现 5 次正面前出现反面次数的方差为多少？
3. 考虑习题 2 中抛掷硬币的试验序列.
   a. 出现 5 次正面前期望投掷次数为多少？
   b. 出现 5 次正面前投掷次数的方差为多少？
4. 设有两个人 A 和 B 试图将篮球投进铁环，在任意一次投掷中，A 成功的概率为 p，他一直投到成功 r 次为止；B 成功的概率为 mp，其中 m 是给定的整数($m=2,3,\cdots$)且 $mp<1$，她一直投到成功 mr 次为止.
   a. 哪个人投掷的期望次数较小？
   b. 哪个人投掷次数的方差较小？
5. 设随机变量 $X_1,\cdots,X_k$ 相互独立，且 X_i 服从参数为 r_i 和 $p(i=1,\cdots,k)$ 的负二项分布. 证明：$X_1+\cdots+X_r$ 服从参数为 $r=r_1+\cdots+r_k$ 和 p 的负二项分布.
6. 设随机变量 X 服从参数为 p 的几何分布，试求 X 的取值为偶数 $0,2,4,\cdots$ 的概率.
7. 设随机变量 X 服从参数为 p 的几何分布. 证明：对于任意非负整数 k，$\Pr(X\geqslant k)=q^k$.
8. 证明(5.5.10)式.
9. 设一个电子系统包含 n 个元件，它们工作与否是相互独立的. 假定这些元件是串联的，正如 3.7 节的习题 5 所描述的. 设每个元件正常工作一段时间后都会坏掉. 最后，假定对于 $i=1,\cdots,n$，每个元件 i 正常工作的周期数是一个离散型随机变量，且服从参数为 p_i 的几何分布. 试确定整个系统正常工作的周期数的分布.

267

10. 设 $f(x|r,p)$ 表示参数为 r 和 p 的负二项分布的概率函数. 如(5.4.1)式所定义的，$f(x|\lambda)$ 表示均值为 λ 的泊松分布的概率函数. 令 $r\to\infty$，$q\to 0$，且在整个过程中，rq 保持为常数 λ 不变. 证明：对于任意固定的非负整数 x，有
$$f(x|r,p)\to f(x|\lambda).$$

11. 证明负二项分布的概率函数也可写成下列形式：

$$f(x|r,p)=\begin{cases}\binom{-r}{x}p^r(-q)^x & 对于\ x=0,1,2,\cdots,\\ 0 & 其他.\end{cases}$$

提示：利用5.3节习题10.

12. 设一台机器生产出有缺陷的零件的概率为 P，但 P 未知．假设 P 具有连续型分布，其概率密度函数为：

$$f(p)=\begin{cases}10(1-p)^9 & 对于\ 0<p<1,\\ 0 & 其他.\end{cases}$$

在给定 $P=p$ 的条件下，设所有的零件是相互独立的．以 X 表示直到第一个有缺陷的零件出现前所观测到的没有缺陷的零件数．如果观测到 $X=12$，求在 $X=12$ 的条件下 P 的概率密度函数.

5.6　正态分布

应用最广泛的连续型分布是正态分布族．这是我们看到的第一个其密度函数的积分不能用解析形式表示的分布，因此，为了计算正态分布的概率值和分位数，需要通过查分布函数表或者通过计算机程序得到.

5.6.1　正态分布的重要性

本节将讨论和定义正态分布，是目前为止在统计中最为重要的概率分布．正态分布之所以有如此重要的地位主要有三条理由：

第一个理由是与正态分布的数学性质有直接联系．在本节及本书的后面几节，我们将说明，如果一个随机样本来自于正态分布，则各种重要的样本观测函数的分布都可以被严格地推导出来，且形式简单．因此，假定随机样本是从正态分布中抽取的，在数学处理上将非常方便.

第二个理由是许多科学家已经观测到在各种物理试验中研究的随机变量通常近似地服从正态分布．例如，正态分布通常可以作为某些随机变量很好的近似分布，如同类人群的身高或体重，玉米杆的长度或重量，老鼠的身长或体重，或是由某种工序生产的钢的韧性．有时观测随机变量的简单变换也服从近似分布.

正态分布非常重要的第三个理由是中心极限定理，在下一节将进行介绍和证明．268如果一个大随机样本来自于某个分布，即使该分布不是近似地服从正态分布，中心极限定理的一个推论是：许多重要的样本观测函数都近似地服从正态分布．尤其是，大随机样本来自于任何具有有限方差的总体分布时，样本均值将近似服从于正态分布．下一节将专门介绍这些内容.

5.6.2 正态分布的性质

1. 分布的定义

如果连续型随机变量 X 的概率密度函数 $f(x|\mu,\sigma^2)(-\infty<\mu<\infty,\sigma>0)$ 具有如下形式：

$$f(x|\mu,\sigma^2)=\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\quad 对于 -\infty<t<\infty. \tag{5.6.1}$$

则称 X 服从均值为 μ、方差为 σ^2 的正态分布.

下面来证明由(5.6.1)式定义的非负函数确实是一个概率密度函数，即证明

$$\int_{-\infty}^{\infty} f(x|\mu,\sigma^2)\,\mathrm{d}x=1. \tag{5.6.2}$$

若令 $y=(x-\mu)/\sigma$，则

$$\int_{-\infty}^{\infty} f(x|\mu,\sigma^2)\,\mathrm{d}x=\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{1/2}}\exp\left(-\frac{1}{2}y^2\right)\mathrm{d}y.$$

令

$$I=\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}y^2\right)\mathrm{d}y. \tag{5.6.3}$$

下面证明 $I=(2\pi)^{1/2}$.

由(5.6.3)式，可得

$$\begin{aligned}I^2=I\cdot I&=\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}y^2\right)\mathrm{d}y\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(y^2+z^2)\right]\mathrm{d}y\,\mathrm{d}z.\end{aligned}$$

令 $y=r\cos\theta$，$z=r\sin\theta$，将积分变量 y 和 z 换成极坐标变量 r 和 θ. 由于 $y^2+z^2=r^2$，

$$I^2=\int_0^{2\pi}\int_0^{\infty}\exp\left(-\frac{1}{2}r^2\right)r\,\mathrm{d}r\,\mathrm{d}\theta=2\pi.$$

因此，$I=(2\pi)^{1/2}$，(5.6.2)式也得到证明.

例 5.6.1 汽车排放物 当汽车发动机燃烧汽油时，排出许多令人不快的污染物. Lorenzen（1980）研究了由 46 辆汽车发动机排出的不同污染物的数量. 其中的一类污染物为氮氧化物. 图 5-1 表示根据这 46 辆汽车发动机排出的氮氧化物数量（按克/英里计算）画出的直方图，且用带有一定均值和方差的正态概率密度函数进行拟合. 直方图中的每个条状面积表示 46 次观测值中介于条状两侧的横轴上两点之间的观测值所占的比例. 例如，第四个条形图（横轴的值从 1.0 至 1.2）面积为 $0.870\times0.2=0.174=8/46$，表示有 8 次观测值介于 1.0 和 1.2 之间. ◀

269

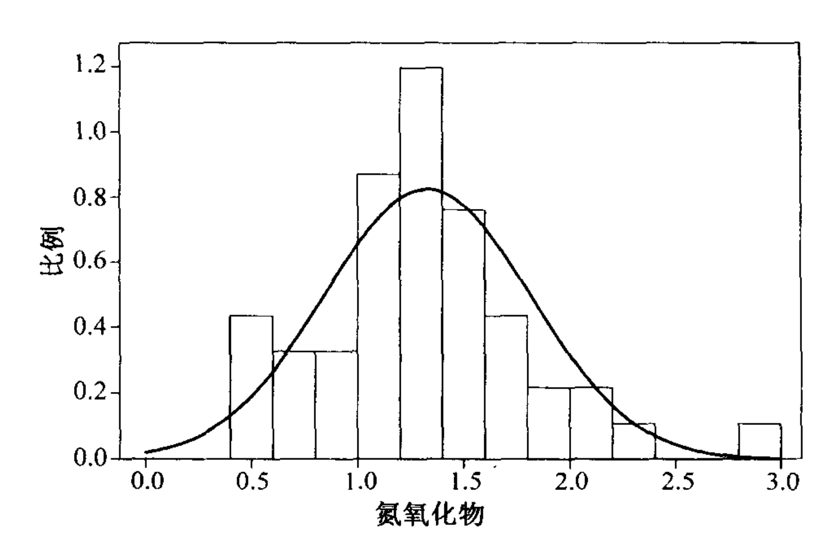


图5-1 习题5.6.1中按普通的驾驶状况，氮氧化物排出量（克/英里）直方图

2. 矩母函数

在正态分布的定义中，参数μ和σ^2为分布的均值和方差. 为了说明这点，必须证明μ确实是由(5.6.1)式定义的概率密度函数的均值，σ^2确实是其方差. 首先来计算正态分布的矩母函数$\psi(t)$，然后再来证明这一事实.

由矩母函数的定义，

$$\psi(t)=\mathrm{E}(\mathrm{e}^{tX})=\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[tx-\frac{(x-\mu)^2}{2\sigma^2}\right]\mathrm{d}x.$$

通过使方括号里的项变成完全平方（见习题24），可以得到关系：

$$tx-\frac{(x-\mu)^2}{2\sigma^2}=\mu t+\frac{1}{2}\sigma^2t^2-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}.$$

因此，

$$\psi(t)=C\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right),$$

270 其中，

$$C=\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{1/2}\sigma}\exp\left\{-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}\right\}\mathrm{d}x.$$

如果以μ代替(5.6.1)式中的$\mu+\sigma^2t$，则由(5.6.2)式知$C=1$. 于是，正态分布的矩母函数为：

$$\psi(t)=\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right)\quad 对于 -\infty<t<\infty. \tag{5.6.4}$$

如果随机变量X服从正态分布，且概率密度函数为(5.6.1)式，则由(5.6.4)式可以得到

$$E(X)=\psi'(0)=\mu$$

和

$$\mathrm{Var}(X)=\psi''(0)-[\psi'(0)]^2=\sigma^2.$$

如此一来，就证明了μ和σ^2确实是由(5.6.1)式定义的正态分布的均值和方差.

既然矩母函数$\psi(t)$对于所有的t都是有限值，那么所有的矩$E(X^k)(k=1,2,\cdots)$也都是有限值.

例5.6.2 股票价格变动 股票价格在长度为u的时间段中变动的一个很常见的模型是，经过时间u后，股票的价格为$S_u=S_0e^{Z_u}$，其中Z_u服从均值为μu、方差为$\sigma^2 u$的正态分布. 在此公式中，S_0是股票当前价格，称σ为股票的波动率. 可以通过Z_u的矩母函数ψ计算得到S_u的期望值：

$$E(S_u)=S_0E(e^{Z_u})=S_0\psi(1)=S_0e^{\mu u+\sigma^2u/2}.$$ ◀

3. 正态分布的形状

由(5.6.1)式可以看到，具有均值为μ、方差为σ^2的正态分布的概率密度函数$f(x|\mu,\sigma^2)$关于点$x=\mu$对称. 因此，μ既是均值又是中位数. 而且，μ还是该分布的众数. 也就是说，概率密度函数$f(x|\mu,\sigma^2)$在点$x=\mu$处达到最大值. 最后，通过对$f(x|\mu,\sigma^2)$求两次导数，可以发现该函数有两个拐点：$x=\mu+\sigma$，$x=\mu-\sigma$.

概率密度函数$f(x|\mu,\sigma^2)$的曲线如图5-2所示，可以看到，该曲线的形状很像一个钟形. 然而每个具有钟形的概率密度函数未必可以用正态分布来近似. 比如，图4-3给出的柯西分布的概率密度曲线是一个对称的钟形图，与图5-2给出的概率密度曲线非常相似. 然而，由于柯西分布的任何矩甚至均值都不存在，所以柯西分布的概率密度曲线的尾部与正态分布的概率密度曲线的尾部有很大差异.

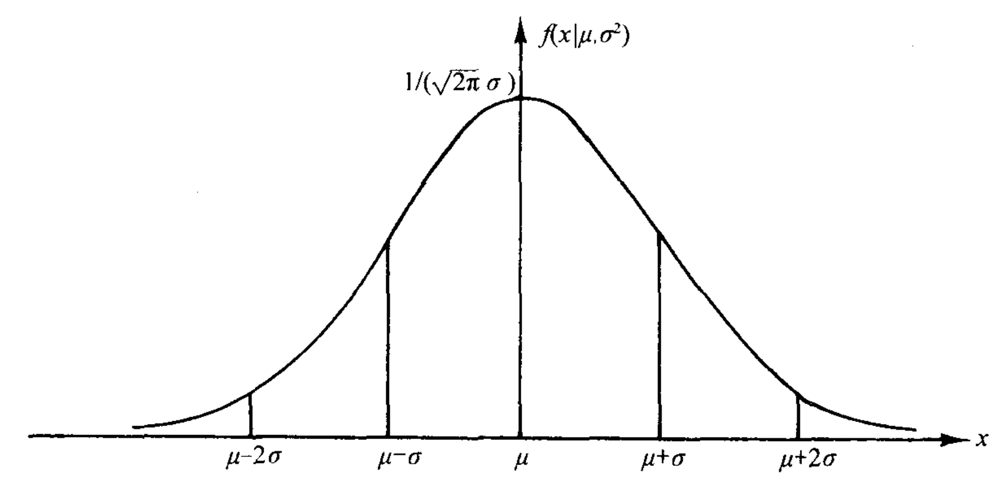


图5-2 正态分布概率密度函数

4. 线性变换

接下来，我们将说明如果随机变量X服从正态分布，则X的任意线性函数也服从正态分布.

定理5.6.1 如果随机变量X服从参数为μ，σ^2的正态分布，若$Y=aX+b$，其中a

和 b 是给定常数，$a\neq0$，则 Y 服从均值为 $a\mu+b$、方差为 $a^2\sigma^2$ 的正态分布.

证明： X 的矩母函数由(5.6.4)式给出，以 ψ_Y 表示 Y 的矩母函数，则

$$\psi_Y(t)=e^{bt}\psi(at)=\exp\left[(a\mu+b)t+\frac{1}{2}a^2\sigma^2t^2\right]\quad 对于 -\infty<t<\infty.$$

将 ψ_Y 的表达式与由(5.6.4)式给出的正态分布的矩母函数进行比较，可以看到 ψ_Y 是均值为 $a\mu+b$、方差为 $a^2\sigma^2$ 的正态分布的矩母函数. ■

5.6.3 标准正态分布

称均值为0、方差为1的正态分布为标准正态分布. 标准正态分布的概率密度函数通常以符号 ϕ 表示，分布函数以 Φ 表示. 这样，

$$\phi(x)=f(x|0,1)=\frac{1}{(2\pi)^{1/2}}\exp\left(-\frac{1}{2}x^2\right)\quad 对于 -\infty<x<\infty \tag{5.6.5}$$

$$\Phi(x)=\int_{-\infty}^{x}\phi(u)\mathrm{d}u\quad 对于 -\infty<x<\infty, \tag{5.6.6}$$

上式中的 u 表示积分中的名义变量.

分布函数 $\Phi(x)$ 不能用基本函数显式的形式表达出来. 因此，标准正态分布或者其他正态分布的概率只能通过近似计算或者利用本书后面的附录表给出的 $\Phi(x)$ 值得到. 表中，只给出 $x\geq0$ 时，$\Phi(x)$ 的值. 由于标准正态分布关于 $x=0$ 对称，则对于每个 $x(-\infty<x<\infty)$，有 $\Pr(X\leq x)=\Pr(X\geq -x)$. 又因为 $\Pr(X\leq x)=\Phi(x)$，$\Pr(X\geq -x)=1-\Phi(-x)$，于是，

272

$$\Phi(x)+\Phi(-x)=1\quad 对于 -\infty<x<\infty. \tag{5.6.7}$$

许多计算机统计分析的软件包都含有一些函数，用于计算标准正态分布的分布函数和分位数函数.

由定理5.6.1知，如果随机变量 X 服从均值为 μ、方差为 σ^2 的正态分布，则随机变量 $Z=(X-\mu)/\sigma$ 服从标准正态分布. 因此，具有任意给定均值和方差的正态分布的概率均可以通过标准正态分布表得到.

例5.6.3 计算正态分布概率 设 X 服从均值为5、标准差为2的正态分布. 现在来确定概率 $\Pr(1<X<8)$ 的值.

令 $Z=(X-5)/2$，则 X 具有标准正态分布，且

$$\Pr(1<X<8)=\Pr\left(\frac{1-5}{2}<\frac{X-5}{2}<\frac{8-5}{2}\right)=\Pr(-2<Z<1.5).$$

进一步，

$$\Pr(-2<Z<1.5)=\Pr(Z<1.5)-\Pr(Z\leq -2)$$
$$=\Phi(1.5)-\Phi(-2)=\Phi(1.5)-[1-\Phi(2)].$$

根据本书后面的附录表，可以查得 $\Phi(1.5)=0.9332$，$\Phi(2)=0.9773$. 于是，

$$\Pr(1<X<8)=0.9105.$$

◀

例 5.6.4 正态分布的分位数 设工程师收集到例 5.6.1 中汽车排放物的数据，他们对于大部分的发动机是否产生严重的污染感兴趣. 比如说，他们可以计算排放物分布的 0.05 分位数，并断言在被检测的该种类型的发动机中，有 95% 超过这一分位数. 以 X 表示某种类型的发动机每英里排放出的氧化物平均克数. 则工程师以正态分布来拟合 X. 图 5-1 给出的正态分布中，均值为 1.329，方差为 0.4844. 则 X 的分布函数 $F(x)=\Phi([x-1.329]/0.4844)$，分位数函数为 $F^{-1}(p)=1.329+0.4844\Phi^{-1}(p)$，$\Phi^{-1}$ 是标准正态分布的分位数函数，它可以通过计算机或者查表得到. 为了从 Φ 表中找到 $\Phi^{-1}(p)$，需要在 $\Phi(x)$ 列找到与 p 最接近的值，然后从 x 列中找到对应的值. 由于表中仅有 $p>0.5$ 的值，根据 $\Phi(-x)=1-\Phi(x)$，可以推出 $\Phi^{-1}(1-p)=-\Phi^{-1}(p)$. 因此，在 $\Phi(x)$ 列查找 0.95（在 0.9495 和 0.9505 之间）所对应的 x 值，可以得到 $x=1.645$（在 1.64 和 1.65 之间），于是，推断 $\Phi^{-1}(0.05)=-1.645$. 所以，X 的 0.05 分位数为 $1.329+0.4844\times(-1.645)=0.5322$. ◀

273

5.6.4 正态分布的比较

图 5-3 给出了具有固定的 μ、不同的 $\sigma(\sigma=1/2,1,2)$ 的正态分布的概率密度曲线. 从图中可以看出，具有较小 σ 值的正态分布概率密度曲线有较高的峰，并且在均值 μ 的两侧非常集中. 然而具有较大 σ 值的正态分布的概率密度曲线相对平坦且比较平缓地延伸在整条直线上.

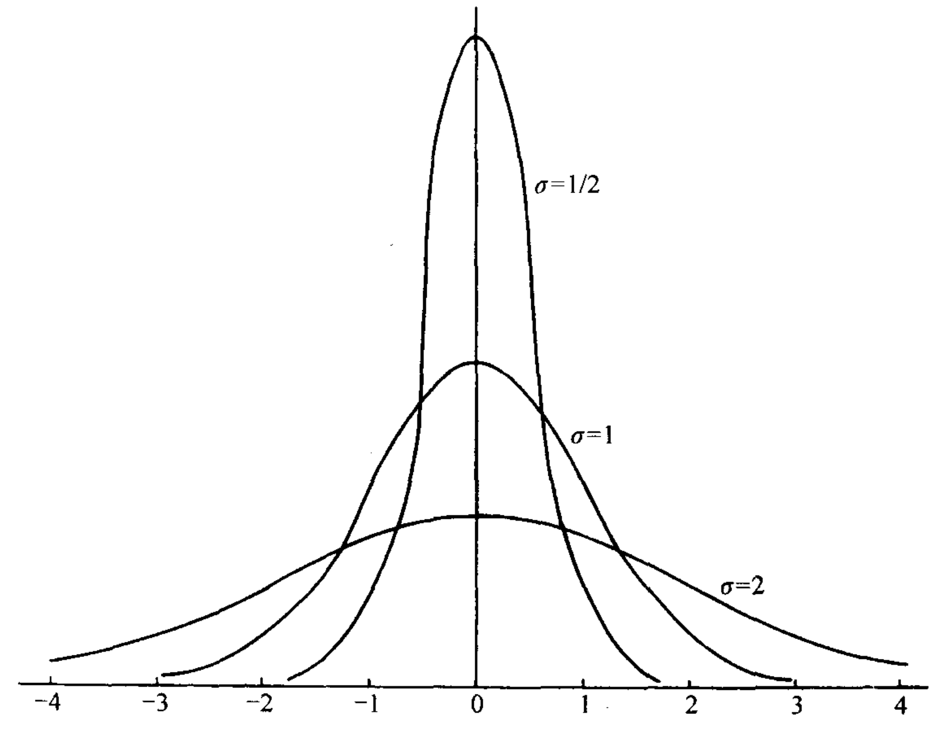


图 5-3 $\mu=0$，$\sigma=1/2$，1，2 的正态分布概率密度曲线

一个重要的事实是，每个正态分布在均值附近的一个标准差内具有相同的概率值，在均值附近的两个标准差内也具有相同的概率值，及在均值附近任何其他固定的每个标准差内都具有相同的概率值. 更为一般地，如果 X 服从均值为 μ、方差为 σ^2 的正态分布，Z 服从标准正态分布，则对于 $k>0$，

$$p_k = \Pr(|x-\mu| \leqslant k\sigma) = \Pr(|Z| \leqslant k).$$

表5-1给出了不同的 k 值对应的概率 p_k. 通过 Φ 表或者计算机程序可以得到这些概率值. 虽然正态分布的概率密度函数在整个实轴上是正的，但可以看到在均值两侧的四个标准差外的总概率值却是很小的，只有0.000 06.

274

表5-1 位于均值 k 个标准差内的标准正态随机变量的概率

k	1	2	3	4	5	10
p_k	0.682 6	0.954 4	0.997 4	0.999 94	$1-6\times10^{-7}$	$1-2\times10^{-23}$

5.6.5 正态随机变量的线性组合

在接下来的定理和推论中，我们将证明下述重要结论：独立并且服从正态分布的随机变量的线性组合仍具有正态分布.

定理5.6.2 *如果随机变量 $X_1,\cdots,X_k$ 相互独立，且 X_i 服从均值为 μ_i、方差为 $\sigma_i(i=1,\cdots,k)$ 的正态分布，则 $X_1+\cdots+X_k$ 服从均值为 $\mu_1+\cdots+\mu_k$、方差为 $\sigma_1^2+\cdots+\sigma_k^2$ 的正态分布.*

证明： 以 $\psi_i(t)$ 表示 $X_i(i=1,\cdots,k)$ 的矩母函数，$\psi(t)$ 表示 $X_1+\cdots+X_k$ 的矩母函数. 由于随机变量 $X_1,\cdots,X_k$ 相互独立，则

$$\psi(t) = \prod_{i=1}^{k} \psi_i(t) = \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right)$$

$$= \exp\left[\left(\sum_{i=1}^{k}\mu_i\right)t + \frac{1}{2}\left(\sum_{i=1}^{k}\sigma_i^2\right)t^2\right] \quad 对于 -\infty < t < \infty.$$

由(5.6.4)式，矩母函数 $\psi(t)$ 是均值为 $\sum_{i=1}^{k}\mu_i$、方差为 $\sum_{i=1}^{k}\sigma_i^2$ 的正态分布的矩母函数. 因此，$X_1+\cdots+X_k$ 的分布是上述定理中给出的分布. ■

下面的推论是由定理5.6.1和定理5.6.2组合得到的.

推论5.6.1 如果随机变量 $X_1,\cdots,X_k$ 相互独立，且 X_i 服从均值为 μ_i、方差为 $\sigma_i(i=1,\cdots,k)$ 的正态分布. 如果 $a_1,\cdots,a_k$ 和 b 为常数，且 $a_1,\cdots,a_k$ 中至少有一个不等于零，则随机变量 $a_1X_1+\cdots+a_kX_k+b$ 服从均值为 $a_1\mu_1+\cdots+a_k\mu_k+b$、方差为 $a_1^2\sigma_1^2+\cdots+a_k^2\sigma_k^2$ 的正态分布. ■

275

据此，也很容易导出来自正态总体随机样本的样本均值的分布.

推论5.6.2 设随机变量 $X_1,\cdots,X_n$ 为来自均值为 μ、方差为 σ^2 的正态总体. 以 $\overline{X}_n$ 表示样本均值，则 $\overline{X}_n$ 服从均值为 μ、方差为 σ^2/n 的正态分布.

证明： 由于 $\overline{X}_n=(1/n)\sum_{i=1}^{n}X_i$，则由推论 5.6.1 可知，$\overline{X}_n$ 的分布是正态分布. 由 4.8 节可知，$E(\overline{X}_n)=\mu$，$Var(\overline{X}_n)=\sigma^2/n$. ■

例 5.6.5 确定样本容量 设一个样本容量为 n 的随机样本来自于均值为 μ、方差为 9 的正态总体. 下面来确定最小值 n，使得

$$\Pr(|\overline{X}_n-\mu|\leqslant 1)\geqslant 0.95.$$

由推论 5.6.2 知，样本均值 $\overline{X}_n$ 具有均值为 μ、标准差为 $3/n^{1/2}$ 的正态分布. 因此，若令

$$Z=\frac{n^{1/2}}{3}(\overline{X}_n-\mu),$$

则 Z 服从标准正态分布. 此例中，n 必须满足

$$\Pr(|\overline{X}_n-\mu|\leqslant 1)=\Pr\left(|Z|\leqslant\frac{n^{1/2}}{3}\right)\geqslant 0.95. \tag{5.6.8}$$

对于每个正数 x，$\Pr(|Z|\leqslant x)\geqslant 0.95$ 当且仅当 $1-\Phi(x)=\Pr(Z>x)\leqslant 0.025$. 从本书后面的标准正态分布表可知，$1-\Phi(x)\leqslant 0.025$ 当且仅当 $x\geqslant 1.96$. 因此，(5.6.8)式中的不等式成立当且仅当

$$\frac{n^{1/2}}{3}\geqslant 1.96.$$

由于可允许的最小的 n 值为 34.6，所以满足题中不等式的样本容量至少为 35. ——◀

例 5.6.6 男人和女人的身高 设在某特定总体中，以英寸计算，女人的身高服从均值为 65、标准差为 1 的正态分布；男人的身高服从均值为 68、标准差为 2 的正态分布. 假定独立地随机选择一个女人和一个男人，下面来确定女人比男人高的概率.

以 W 表示被选中的女人的身高，M 表示随机被选中的男人的身高. 则差 $W-M$ 服从均值为 $65-68=-3$ 和方差为 $1^2+2^2=5$ 的正态分布. 因此，若令

$$Z=\frac{1}{5^{1/2}}(W-M+3),$$

则 Z 服从标准正态分布. 从而有

276

$$\begin{aligned}\Pr(W>M)&=\Pr(W-M>0)\\&=\Pr\left(Z>\frac{3}{5^{1/2}}\right)=\Pr(Z>1.342)\\&=1-\Phi(1.342)=0.090.\end{aligned}$$

于是，女人比男人高的概率为 0.090. ——◀

5.6.6 对数正态分布

用正态分布拟合随机变量的对数非常普遍. 由于这个原因，在作对数变换之前的随机变量的分布有一个特定的名称. 如果 $\ln(X)$ 服从均值为 μ、方差为 σ^2 的正态分布，则称 X 服从参数为 μ 和 σ^2 的对数正态分布.

例5.6.7 滚珠轴承的失效时间 有些产品容易磨损和破裂，所以为了估算其有效使用寿命，经常对它们的耐磨性进行检测. Lawless(1982，例5.2.2)描述 Lieblein 和 Zelen(1956)取得的数据进行了描绘，该数据是对23个滚珠轴承进行旋转并观测其在破裂前成百万计的转数. 对数正态分布是一个拟合失效时间非常有用的模型. 图5-4给出23个滚珠轴承使用寿命的直方图，及根据该数据拟合出的具有确定参数的对数正态分布的概率密度曲线. 图5-4中直方图条形的解释与例5.6.1描述的图5-1有相同的解释. 假定工程师感兴趣的是想知道一个滚珠轴承直到有90%的机会失效需要等待多长时间，则只要得到寿命分布的90%分位数. 以 X 表示一个滚珠轴承到失效时经历的时间. 图5-4描绘的是 X 的对数正态分布具有参数4.15和 0.5334^2. 则 X 的分布函数为 $F(x)=\Phi([\ln(x)-4.15]/0.5334)$，分位数函数为 $F^{-1}(p)=e^{4.15+0.5334\Phi^{-1}(p)}$，其中 Φ^{-1} 表示标准正态分布的分位数函数. $p=0.9$ 时，$\Phi^{-1}(0.9)=1.28, F^{-1}(0.9)=125.6$. ◀

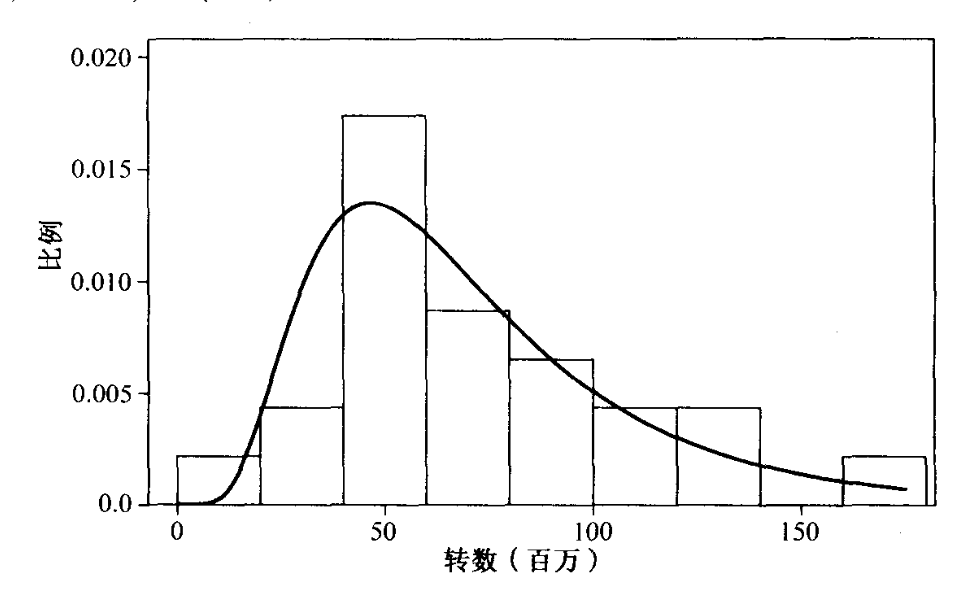


图5-4 例5.6.7中滚珠轴承寿命的直方图及拟合的对数正态概率密度曲线

277

根据正态分布的矩母函数，很容易得到对数正态随机变量的矩. 若 $Y=\ln(X)$ 服从均值为 μ、方差为 σ^2 的正态分布，则 Y 的矩母函数为 $\psi(t)=\exp(\mu t+0.5\sigma^2t^2)$. 但是，$\psi$ 的定义为 $\psi(t)=E(e^{tY})$. 由于 $Y=\ln(X)$，有

$$\psi(t)=E(e^{tY})=E(e^{t\ln(X)})=E(X^t).$$

对于所有的实数 t 有 $E(X^t)=\psi(t)$，尤其是 X 的均值和方差为

$$E(X)=\psi(1)=\exp(\mu+0.5\sigma^2), \tag{5.6.9}$$
$$\mathrm{Var}(X)=\psi(2)-\psi(1)^2=\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1].$$

例5.6.8 股票和期权价格 考虑一种如例5.6.2描述的股票，它当前的价格为 S_0，经过时间 u 后，股票的价格为 $S_u=S_0e^{Z_u}$，其中 Z_u 服从均值为 μu、方差为 σ^2u 的正态分布. 注意到 $S_0e^{Z_u}=e^{Z_u+\ln(S_0)}$ 及 $Z_u+\ln(S_0)$ 服从均值为 $\mu u+\ln(S_0)$、方差为 σ^2u 的

正态分布，于是，S_u 服从参数为 $\mu u+\ln(S_0)$ 和 $\sigma^2 u$ 的对数正态分布.

Black and Scholes (1973) 在假定股票价格服从对数正态分布的条件下得到股票期权定价的公式. 在本例接下来的部分，考虑单一时段 u，且将股票的价格写为 $S_u=S_0 e^{\mu u+\sigma u^{1/2}Z}$，其中 Z 服从标准正态分布. 假设我们要对在将来某个特定时间 u 以确定价格 q 购买一份该股票的期权定价. 如例 4.1.4，我们使用风险中性定价法. 即，使 $E(S_u)$ 的现值等于 S_0. 如果 u 以年为单位计量，每年的无风险利率为 r，则 $E(S_u)$ 的现值为 $e^{-ru}E(S_u)$.（这里假定是连续复利，而不是像例 4.1.4 的单利，将在习题 25 来检验连续复利的影响.）但是 $E(S_u)=S_0 e^{\mu u+\sigma^2 u/2}$. 在作风险中性定价时令 $S_0=e^{-ru}S_0 e^{\mu u+\sigma^2 u/2}$，得到 $\mu=r-\sigma^2/2$.

现在来确定特定期权的价格. 在时刻 u 处，期权价值为 $h(S_u)$，其中

$$h(s)=\begin{cases}s-q & \text{对于 } s>q,\\ 0 & \text{否则.}\end{cases}$$

现在，令 $\mu=r-\sigma^2/2$，容易看到 $h(S_u)>0$ 当且仅当

$$Z>\frac{\ln\left(\frac{q}{S_0}\right)-(r-\sigma^2/2)u}{\sigma u^{1/2}}. \tag{5.6.10}$$

接下来，把(5.6.10)式中右边视为常数 c. 期权的风险中性价格是 $E(h(S_u))$ 的现值，它等于

$$e^{-ru}E[h(S_u)]=e^{-ru}\int_c^\infty [S_0 e^{[r-\sigma^2/2]u+\sigma u^{1/2}z}-q]\frac{1}{(2\pi)^{1/2}}e^{-z^2/2}dz. \tag{5.6.11}$$

278

为了计算(5.6.11)式的积分，需在 $-q$ 处把被积函数分成两部分. 第二部分积分可看成正态分布概率密度函数的常数倍，也就是说

$$-e^{-ru}q\int_c^\infty \frac{1}{(2\pi)^{1/2}}e^{-z^2/2}dz=-e^{-ru}q[1-\Phi(c)].$$

(5.6.11)式中第一部分积分为

$$e^{-\sigma^2 u/2}S_0\int_c^\infty \frac{1}{(2\pi)^{1/2}}e^{-z^2/2+\sigma u^{1/2}z}dz.$$

可以通过把该项配方的方法变成正态概率密度函数在一定区间上积分的常数倍，取完全平方后的结果为：

$$e^{-\sigma^2 u/2}S_0\int_c^\infty \frac{1}{(2\pi)^{1/2}}e^{-(z-\sigma u^{1/2})^2/2+\sigma^2 u/2}dz=S_0[1-\Phi(c-\sigma u^{1/2})].$$

最后，将两部分积分合起来，成为期权价格，并利用 $1-\Phi(x)=\Phi(-x)$，得：

$$S_0\Phi(\sigma u^{1/2}-c)-qe^{-ru}\Phi(-c). \tag{5.6.12}$$

这是著名的期权定价 Black-Scholes 公式. 举个简单的例子，设 $q=S_0$，$r=0.06$（6%的利率），$u=1$（一年期），$\sigma=0.1$. 由(5.6.12)式可知，期权价格为 $0.0746S_0$. 如果 S_u 的分布不同于此处所用的，则可用模拟的技巧（见第 11 章）来计算期权价格. ──◀

在本节习题17可以求得对数正态分布的概率密度函数. 由标准正态分布的分布函数$\Phi(x)$很容易得到对数正态分布的分布函数. 设X服从参数为μ和σ^2的对数正态分布. 则

$$\Pr(X \leqslant x) = \Pr(\ln(X) \leqslant \ln(x)) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right).$$

本节前面正态随机变量线性组合的结果可以化为对数正态随机变量幂的乘积的结果. 独立正态随机变量和的结果可以化为独立对数正态变量的乘积的结果.

5.6.7 小结

本节介绍了正态分布族. 每个正态分布的参数是分布的均值和方差. 独立正态随机变量的线性组合仍服从正态分布，其均值为各个均值的线性组合，方差由推论4.3.1所决定. 特别地，如果随机变量X服从均值为μ、方差为σ^2的正态分布，则$(X-\mu)/\sigma$服从标准正态分布(均值为0,方差为1). 可以通过查表或者计算机程序得到标准正态分布的概率和分位数，而一般正态分布的概率和分位数需要通过标准正态分布的概率和分位数得到. 比如若X服从均值为μ、方差为σ^2的正态分布，则分布函数为$F(x) = \Phi([x-\mu]/\sigma)$，$X$的分位数为$F^{-1}(p) = \mu + \Phi^{-1}(p)\sigma$，其中$\Phi$为标准正态分布的分布函数.

279

习题

1. 找出标准正态分布的0.5，0.25，0.75，0.1，0.9的分位数.
2. 设X服从均值为1、方差为4的正态分布，试求下列概率值：
   a. $\Pr(X \leqslant 3)$ b. $\Pr(X > 1.5)$ c. $\Pr(X = 1)$ d. $\Pr(2 < X < 5)$
   e. $\Pr(X \geqslant 0)$ f. $\Pr(-1 < X < 0.5)$ g. $\Pr(|X| \leqslant 2)$ h. $\Pr(1 \leqslant -2X + 3 \leqslant 8)$.
3. 如果在某一特定地点的温度（以华氏度计量）服从正态分布，其均值为68度，标准差为4度. 试求：当以摄氏度计量时在同一地点温度的分布.
4. 在习题3中，计算以华氏度计量的温度的0.25和0.75分位数.
5. 设X_1、X_2、X_3为存储器芯片寿命，它们是相互独立的. 假定X_i服从均值为300小时、标准差为10小时的正态分布. 计算至少有一个存储器芯片的寿命不小于290小时的概率.
6. 如果随机变量X的矩母函数为$\psi(t) = e^{t^2}$，对于$-\infty < t < \infty$. 求X的分布.
7. 设在某电路中测定的电压服从均值为120、标准差为2的正态分布. 若进行三次独立的电压测量，则三次测量值都在116和118之间的概率为多少?
8. 计算积分$\int_0^\infty e^{-3x^2}dx$.
9. 一直杆由A，B，C三节连接而成，每一节由不同的机器生产. 以英寸计算，A节的长度服从均值为20、方差为0.04的正态分布；B节的长度服从均值为14、方差为0.01的正态分布；C节的长度服从均值为26、方差为0.04的正态分布. 如图5-5所示，将三节连接起来，每两节之间都有2英寸重叠. 如果该杆的总长度(以英寸计算)在55.7至56.3之间，则可用于建造机翼. 计算该杆可被利用的概率.

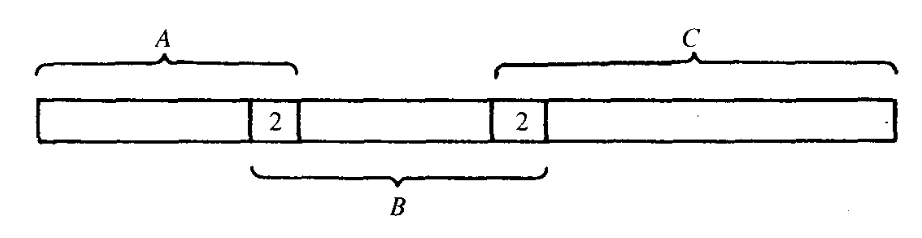


图 5-5 习题 9 中直杆的分段

10. 样本容量为 25 的随机样本来自于均值为 μ、标准差为 2 的正态总体. 样本均值在 μ 的一个单位邻域内的概率为多少?

11. 设一样本容量为 n 的随机样本来自于均值为 μ、标准差为 2 的正态总体. 求使下述不等式成立的最小的 n 值:

$$\Pr(|\overline{X}_n-\mu|<0.1)\geqslant 0.9.$$

12. a. 根据本书后面的附录表中给出的标准正态分布表画出分布函数 Φ 的图形.

    b. 根据 a. 画出的正态分布函数图, 画出均值为 -2、标准差为 3 的正态分布函数图.

13. 设一个大箱子里螺丝钉的直径服从均值为 2 厘米、标准差为 0.03 厘米的正态分布. 假定另一个大箱子里螺丝帽上洞的直径服从均值为 2.02 厘米、标准差为 0.04 厘米的正态分布. 如果螺丝帽上洞的直径比螺丝钉的直径大但又不大于 0.05 厘米, 则螺丝和螺丝帽可以互相匹配. 如果随机选取一个螺丝钉和一个螺丝帽, 求它们可以匹配的概率为多少? [280]

14. 设在某次高等数学考试中, 来自 A 大学的学生考试分数服从均值为 625、方差为 100 的正态分布, 来自 B 大学的学生考试分数服从均值为 600、方差为 150 的正态分布. 如果有 2 个 A 大学的学生、3 个 B 大学的学生参加该次考试, 则来自 A 大学的 2 个学生的平均成绩高于来自 B 大学的 3 个学生的平均成绩的概率为多少? 提示: 确定两个平均成绩差的分布.

15. 假定在某特定人群中, 有 10% 的人患有青光眼疾病. 对于患有青光眼疾病的人来说, 其眼压的测量值服从均值为 25、方差为 1 的正态分布. 对未患有青光眼疾病的人来说, 其眼压的测量值服从均值为 20、方差为 1 的正态分布. 假定从该人群中随机选取一个人, 并测量其眼压.

    a. 在给定 $X=x$ 的条件下, 确定此人患有青光眼疾病的条件概率.

    b. x 为何值时, a. 中的条件概率大于 1/2?

16. 设两随机变量 X 和 Y 的联合概率密度函数为

$$f(x,y)=\frac{1}{2\pi}e^{-(1/2)(x^2+y^2)}\quad -\infty<x<\infty \text{ 且 } -\infty<y<\infty.$$

    求概率 $\Pr(-\sqrt{2}<X+Y<2\sqrt{2})$ 的值.

17. 设一随机变量 X 服从参数为 μ 和 σ^2 的对数正态分布. 求 X 的概率密度函数.

18. 设随机变量 X 和 Y 相互独立, 且均服从标准正态分布. 证明: X/Y 服从柯西分布.

19. 假定某系统装置压力的测量值 X 服从均值为 μ、方差为 1 的正态分布, 其中 μ 为真实压力值. 设真实压力值 μ 未知, 但在区间 $[5,15]$ 上服从均匀分布. 若观测到 $X=8$, 求在给定 $X=8$ 的条件下, μ 的概率密度函数.

20. 设 X 服从参数为 3 和 1.44 的对数正态分布. 求 $X\leqslant 6.05$ 的概率值.

21. 设随机变量 X 和 Y 相互独立, 且 $\ln X$ 服从均值为 1.6、方差为 4.5 的正态分布, $\ln Y$ 服从

均值为3、方差为6的正态分布，求乘积XY的分布.

22. 设X服从参数为μ和σ^2的对数正态分布，求$1/X$的分布.
23. 设X服从参数为4.1和8的对数正态分布，求$3X^{1/2}$的分布.
24. 本书中，配方的方法已经用过几次. 这种方法是把二次线性多项式变成一个完全平方项加上一个常数. 证明下述等式，该等式是配方的一般形式：

$$\sum_{i=1}^{n} a_i(x-b_i)^2+cx=\left(\sum_{i=1}^{n} a_i\right)\left(x-\frac{\sum_{i=1}^{n} a_ib_i-c/2}{\sum_{i=1}^{n} a_i}\right)^2+\sum_{i=1}^{n} a_i\left(b_i-\frac{\sum_{i=1}^{n} a_ib_i}{\sum_{i=1}^{n} a_i}\right)^2+\left(\sum_{i=1}^{n} a_i\right)^{-1}\left[c\sum_{i=1}^{n} a_ib_i-c^2/4\right],$$

假设$\sum_{i=1}^{n} a_i\neq 0$.

25. 例5.6.8中，考虑连续复利的影响. 假定S_0美元连续u年获得每年利率为r的连续复利. 证明：在u年末，本金加利息的值为$S_0\mathrm{e}^{ru}$. 提示：设利率在每个长度为u/n年的区间计算复利n次. 在每n个时间区间末，本金被扩大$1+ru/n$倍. 令$n\to\infty$，便得到结果.

281

5.7　中心极限定理

设总体具有均值μ和有限方差σ^2，则来自于该总体的随机大样本的样本均值近似地服从均值为μ、方差为σ^2/n的正态分布. 利用此结果可以证明为什么可以用正态分布来刻划许多由大量独立部分构成的随机变量. 中心极限定理的另一种变形是应用在随机变量不是相互独立的情形. 本节还将介绍δ方法，该方法可以用来求随机变量函数的近似分布.

5.7.1　定理的陈述

在5.6节中已经证明过：如果样本容量为n的一个样本来自于均值为μ、方差为σ^2的正态总体，则样本均值服从均值为μ、方差为σ^2/n的正态分布. 在本节，我们将讨论中心极限定理，可以粗略地叙述该定理的一个简单形式：从服从均值为μ、方差为σ^2的任意分布的总体中随机抽取一个容量为n的随机样本，则样本的均值近似服从均值为μ、方差为σ^2/n的正态分布.

在18世纪初期，棣莫佛对于从伯努利分布的总体中抽取的随机样本，得到该结论. 对于来自于任意总体中的随机样本，该结论由林德伯格和莱维在20世纪20年代得到. 接下来将会给出该定理的精确叙述，在本节的最后部分也将给出该定理证明的概要. 此外，还将介绍有关分布不一定相同但相互独立的随机变量和的中心极限

定理. 同时，会给出一些例子说明两个定理的应用.

1. 关于样本均值的林德伯格－莱维中心极限定理

正如5.6节所介绍的，以 Φ 表示标准正态分布的分布函数.

定理5.7.1 如果随机变量 $X_1,\cdots,X_n$ 是来自于均值为 μ、方差为 $\sigma^2(0<\sigma^2<\infty)$ 的一个给定总体容量为 n 的一个随机样本，则对于每个固定的 x，

$$\lim_{n\to\infty}\Pr\left[\frac{n^{1/2}(\overline{X}_n-\mu)}{\sigma}\leqslant x\right]=\Phi(x). \qquad (5.7.1)\blacksquare$$

将在本节后面给出该定理证明的概要.

(5.7.1)式的解释如下：如果一个大样本来自于服从均值为 μ、方差为 σ^2 的任意分布的总体中，无论该分布是离散的还是连续的，都有：随机变量 $n^{1/2}(\overline{X}_n-\mu)/\sigma$ 近似服从标准正态分布. 因此，$\overline{X}_n$ 近似服从均值为 μ、方差为 σ^2/n 的正态分布，即 $\sum_{i=1}^{n}X_i$ 近似服从均值为 $n\mu$、方差为 $n\sigma^2$ 的正态分布.

282

例5.7.1 抛掷硬币 假定抛掷一均匀硬币900次，现在来确定出现正面次数多于495次的概率.

对于 $i=1,\cdots,900$，若第 i 次抛掷出现正面，令 $X_i=1$，否则，令 $X_i=0$. 则 $E(X_i)=1/2,\mathrm{Var}(X_i)=1/4$. 因此，$X_1,\cdots,X_{900}$ 是来自于均值为1/2、方差为1/4的分布总体中容量为 $n=900$ 的一个样本. 根据中心极限定理，出现正面的总次数 $H=\sum_{i=1}^{900}X_i$ 近似服从正态分布，其均值为 $(900)(1/2)=450$，方差为 $(900)(1/4)=225$，标准差为 $(225)^{1/2}=15$. 于是，随机变量 $Z=(H-450)/15$ 近似地服从标准正态分布. 则

$$\Pr(H>495)=\Pr\left(\frac{H-450}{15}>\frac{495-450}{15}\right)=\Pr(Z>3)\approx 1-\Phi(3)=0.0013. \qquad ◀$$

例5.7.2 来自于均匀分布的样本 设一容量 $n=12$ 的随机样本来自于[0,1]上的均匀分布总体. 现在来确定 $\Pr\left(\left|\overline{X}_n-\frac{1}{2}\right|\leqslant 0.1\right)$ 的值.

区间[0,1]上均匀分布的均值为1/2，方差为1/12(见4.3节中的习题3). 由于此例中 $n=12$，由中心极限定理知，$\overline{X}_n$ 近似地服从于均值为1/2、方差为1/144的正态分布，则 $Z=12\left(\overline{X}_n-\frac{1}{2}\right)$ 近似地服从于标准正态分布. 因此，

$$\Pr\left(\left|\overline{X}_n-\frac{1}{2}\right|\leqslant 0.1\right)=\Pr\left[12\left|\overline{X}_n-\frac{1}{2}\right|\leqslant 1.2\right]$$

$$=\Pr(|Z|\leqslant 1.2)=2\Phi(1.2)-1=0.7698.$$

计算机曾经通过将12个均匀伪随机数相加再减去6的方法来产生标准正态伪随机数. ◀

2. 中心极限定理的作用

中心极限定理提供了这样一个似乎合理的解释：物理试验中研究的许多随机变量都近似地服从正态分布. 比如说，一个人的身高与许多因素有关，如果每个人的身高由许多单个因素的和所决定，则大量人的身高服从正态分布. 一般地，中心极限定理可以解释为什么许多随机变量的和近似地服从正态分布，即使和中每个随机变量的分布并不是正态的.

283

例 5.7.3 确定模拟次数 在 4.8 节例 4.8.2 中，环境工程师为了估计水中杂质比例，需要确定进行模拟的次数. 在那个例子中，利用切比雪夫不等式估计出需模拟 2 000 000 次才能保证估计出的平均比例偏离真实的平均比例小于 0.005 的概率至少为 0.98. 在此例中，我们利用中心极限定理来确定一个比较小的模拟次数，而且仍然可以保证有同样精度的上界. 平均比例的估计为 n 次模拟得到的比例 $R_1,\cdots,R_n$ 的平均值$\overline{R}_n$. 如例4.8.2 所指出的，每个 R_i 的方差为 $\sigma^2\leqslant 1$. 因此，由中心极限定理，$\overline{R}_n$ 近似服从均值为真实比例 $\mathrm{E}(R_i)$、方差不超过 $1/n$ 的正态分布. 由于$\overline{R}_n$ 接近均值的概率随着方差的增加而较少，可以得到

$$\Pr(|\overline{R}_n-\mathrm{E}(R_i)|<0.005)\approx\Phi\left(\frac{0.005}{\sigma/\sqrt{n}}\right)-\Phi\left(\frac{-0.005}{\sigma/\sqrt{n}}\right)\geqslant\Phi\left(\frac{0.005}{1/\sqrt{n}}\right)-\Phi\left(\frac{-0.005}{1/\sqrt{n}}\right)$$
$$=2\Phi(0.005\sqrt{n})-1.$$

若 $2\Phi(0.005\sqrt{n})-1=0.98$，可以得到

$$n=\frac{1}{0.005^2}\Phi^{-1}(0.01)^2=40\,000\times 2.326^2=216\,411.$$

即，由中心极限定理计算所需的模拟次数，只有切比雪夫不等式所确定的模拟次数的 10% 的比例多一点. （事实上由于 σ^2 不大于 1/4，所以，真正需要的模拟次数为 $n=54\,103$. 见4.7 节的习题 14 的证明：区间[0,1]上离散分布的方差不超过 1/4. 对于连续型分布的证明稍微复杂一点，但该结论也同样成立.） ◀

5.7.2 δ 方法

设随机变量 $X_1,\cdots,X_n$ 是来自于具有有限均值 μ、有限方差 σ^2 的分布总体中的一个随机样本，中心极限定理说明 $n^{1/2}(\overline{X}_n-\mu)/\sigma$ 近似地服从标准正态分布. 现在我们感兴趣的是$\overline{X}_n$ 的函数 α 的分布. 假定 α 是可微函数，它在 μ 点处的导数非零. 通过统计学中著名的 δ 方法可以得到 $\alpha(\overline{X}_n)$ 的近似分布.

由$\overline{X}_n$ 的近似分布可知，对于大样本，$\overline{X}_n$ 以非常大的概率接近于 μ. 因此，$\overline{X}_n-\mu$ 很小的概率非常大. 于是，将函数 $\alpha(\overline{X}_n)$ 看作是级数在 μ 处的展开式，忽视 $(\overline{X}_n-\mu)^2$ 和更高次数的项. 如此一来，有

$$\alpha(\overline{X}_n)\approx\alpha(\mu)+\alpha'(\mu)(\overline{X}_n-\mu).$$

且

$$\frac{n^{1/2}}{\sigma}\frac{1}{\alpha'(\mu)}[\alpha(\overline{X}_n)-\alpha(\mu)]\approx\frac{n^{1/2}}{\sigma}(\overline{X}_n-\mu). \tag{5.7.2}$$

284

可以推断，(5.7.2)式左边近似地服从标准正态分布，这是因为右边近似地服从标准正态分布. 换句话说，对于大样本，$\alpha(\overline{X}_n)$近似服从均值为$\alpha(\mu)$、方差为$\sigma^2[\alpha'(\mu)]^2/n$的正态分布.

例 5.7.4 泊松随机变量 设随机变量$X_1,\cdots,X_n$是来自于均值为θ的泊松分布总体的一个样本. 以$\overline{X}_n$表示样本均值. 则$\mu=\theta$，$\sigma^2=\theta$. 由中心极限定理，$n^{1/2}(\overline{X}_n-\theta)/\theta^{1/2}$近似地服从标准正态分布. 现考虑函数$\alpha(x)=1/x$，我们来求$\alpha(\overline{X}_n)$的近似分布. 对$\alpha(x)$求导，得到$\alpha'(x)=-1/x^2$. 由$\delta$方法知，$1/\overline{X}_n$近似地服从正态分布，其均值为$1/\theta$，方差为$\theta/[n\theta^4]=(n\theta^3)^{-1}$. ◀

在例 5.7.4 中，中心极限定理说明$\overline{X}_n$以非常大的概率接近于θ. 由标准正态分布的分布函数可以近似得到$|\overline{X}_n-\theta|$小于某个很小的常数c的概率：

$$\Pr\left(|\overline{X}_n-\theta|<c\right)\approx2\Phi\left(cn^{1/2}\theta^{-1/2}\right)-1. \tag{5.7.3}$$

在大多数统计应用中，假定θ是未知的. 在这种情况下，(5.7.3)式的概率无法计算，因为$\overline{X}_n$的近似方差依赖于θ. 由于此原因，我们希望通过一个函数α对$\overline{X}_n$进行变换，使得$\alpha(\overline{X}_n)$的方差已知. 称这样的函数为*方差稳定化变换*. 反过来，经常可以通过δ方法得到方差稳定化变换. 一般地，注意到$\alpha(\overline{X}_n)$的近似分布有方差$\alpha'(\mu)^2\sigma^2/n$. 为了使该方差变成常数，需要使$\alpha'(\mu)$变为$1/\sigma$的常数倍. 如果σ^2是μ的函数$g(\mu)$，则通过下述等式可以达到要求：

$$\alpha(\mu)=\int_a^{\mu}\frac{\mathrm{d}x}{g(x)^{1/2}}, \tag{5.7.4}$$

其中，a是使得积分有限的任意常数.

例 5.7.5 泊松随机变量 在例 5.7.4 中，有$\sigma^2=\theta=\mu$. 根据(5.7.4)式，令

$$\alpha(\mu)=\int_0^{\mu}\frac{\mathrm{d}x}{x^{1/2}}=2\mu^{1/2}.$$

可以看到，$2\overline{X}_n^{1/2}$近似地服从均值为$2\theta^{1/2}$、方差为$1/n$的正态分布. 对于每个$c>0$，有

$$\Pr\left(|2\overline{X}_n^{1/2}-2\theta^{1/2}|<c\right)\approx2\Phi(cn^{1/2})-1. \tag{5.7.5}$$

在第 7 章中，将看到当θ未知时，如何利用(5.7.5)式中的不等式估计θ. ◀

285

5.7.3 扩展研究：独立随机变量和的中心极限定理（李雅普洛夫）

现在，我们来阐述一个应用在独立但未必同分布的随机变量序列$X_1,X_2,\cdots$情形的中心极限定理. 这个定理是由李雅普洛夫在 1901 年首次证明的. 对于$i=1,\cdots,n$，假定$\mathrm{E}(X_i)=\mu_i,\mathrm{Var}(X_i)=\sigma_i^2$. 令

$$Y_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}}. \tag{5.7.6}$$

则 $E(Y_n)=0, \mathrm{Var}(Y_n)=1$. 下面的定理给出了 Y_n 近似地服从标准正态分布的一个充分条件.

定理 5.7.2 设随机变量 $X_1, X_2, \cdots$ 相互独立，且 $E(|X_i - \mu_i|^3) < \infty, i=1,2,\cdots$. 假定

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left(|X_i - \mu_i|^3\right)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}} = 0. \tag{5.7.7}$$

最后，令 Y_n 为(5.7.6)式中定义的随机变量，则对于每个固定的数 x，有

$$\lim_{n\to\infty} \Pr(Y_n < x) = \Phi(x). \tag{5.7.8}$$ ■

该定理可解释如下：如果(5.7.7)式成立，则对于每个很大的 n，$\sum_{i=1}^{n} X_i$ 近似地服从正态分布，其均值为 $\sum_{i=1}^{n} \mu_i$、方差为 $\sum_{i=1}^{n} \sigma_i^2$. 应该注意到，当 $X_1, X_2, \cdots$ 同分布，且三阶矩存在时，(5.7.7)式自然成立，且(5.7.8)式变为(5.7.1)式.

有必要强调一下林德伯格－莱维定理和李雅普洛夫定理的不同之处. 林德伯格－莱维定理应用于独立同分布的随机变量，只要满足每个随机变量的方差有限即可应用此定理. 而李雅普洛夫定理应用于独立未必同分布的随机变量，必须假定每个随机变量的三阶矩有限且满足(5.7.7)式才能应用此定理.

有关伯努利随机变量的中心极限定理

通过应用李雅普洛夫定理，可以得到如下结果：

286 **定理 5.7.3** 设随机变量 $X_1, \cdots, X_n$ 相互独立，且 X_i 服从参数为 $p_i (i=1,2,\cdots)$ 的伯努利分布. 假定无穷级数 $\sum_{i=1}^{\infty} p_i q_i$ 发散，令

$$Y_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} p_i}{\left(\sum_{i=1}^{n} p_i q_i\right)^{1/2}}. \tag{5.7.9}$$

则对于每个固定的数 x，有

$$\lim_{n\to\infty} \Pr(Y_n \leqslant x) = \Phi(x). \tag{5.7.10}$$

证明： 此处 $\Pr(X_i = 1) = p_i, \Pr(X_i = 0) = q_i$. 因此，$E(X_i) = p_i, \mathrm{Var}(X_i) = p_i q_i$，且

$$E(|X_i - p_i|^3) = p_i q_i^3 + q_i p_i^3 = p_i q_i (p_i^2 + q_i^2) \leqslant p_i q_i. \tag{5.7.11}$$

于是，

$$\frac{\sum_{i=1}^{n} E(|X_i - p_i|^3)}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}} \leqslant \frac{1}{\left(\sum_{i=1}^{n} p_i q_i\right)^{1/2}}. \tag{5.7.12}$$

由于无穷级数 $\sum_{i=1}^{\infty}p_iq_i$ 发散，则当 $n\to\infty$ 时，$\sum_{i=1}^{n}p_iq_i\to\infty$. 由(5.7.12)式可以看出，(5.7.7)式成立. 反过来，由定理5.7.2知，(5.7.8)式一定成立. 由于(5.7.10)式只是(5.7.8)式对于所考虑的特定随机变量的简单重述，所以该定理得证. ■

定理5.7.3表明，如果无穷级数 $\sum_{i=1}^{\infty}p_iq_i$ 发散，则大量独立伯努利随机变量的和 $\sum_{i=1}^{n}X_i$ 近似地服从均值为 $\sum_{i=1}^{n}p_i$、方差为 $\sum_{i=1}^{n}p_iq_i$ 的正态分布. 应该记住，典型的实际问题只涉及有限个随机变量 $X_1,\cdots X_n$，而不是随机变量的无穷序列. 在此问题中，考虑无穷级数 $\sum_{i=1}^{\infty}p_iq_i$ 发散与否是没有任何意义的，因为只须确定有限个值 $p_1,\cdots p_n$. 因此，从某种意义上来说，总是可以用正态分布来近似 $\sum_{i=1}^{n}X_i$. 关键的问题是该正态分布是否是 $\sum_{i=1}^{n}X_i$ 的真实分布的很好近似. 当然，这依赖于 $p_1,\cdots,p_n$.

由于随着 $\sum_{i=1}^{n}p_iq_i\to\infty$，$\sum_{i=1}^{n}X_i$ 越来越接近于正态分布，因此，当 $\sum_{i=1}^{n}p_iq_i$ 很大时，正态分布提供了一个很好的近似. 而且，由于 $p_i=1/2$ 时，p_iq_i 取得最大值，所以，当 n 很大，$p_1,\cdots,p_n$ 接近于1/2时，近似的程度最好.

例5.7.6 考试问题 设一门考试共有99个问题，按从易到难的顺序排列起来. 假定某个学生第一个问题回答正确的概率为0.99；第二个问题回答正确的概率为0.98；一般地，第 i 个问题回答正确的概率为 $1-i/100, i=1,\cdots,99$. 设所有问题的回答是相互独立的，至少回答正确60个问题才能通过考试. 现在来确定一下该学生通过考试的概率.

287

若第 i 个问题回答正确，令 $X_i=1$，否则，令 $X_i=0$，则 $\mathrm{E}(X_i)=p_i=1-(i/100)$，$\mathrm{Var}(X_i)=p_iq_i=(i/100)[1-(i/100)]$. 而且，

$$\sum_{i=1}^{99}p_i=99-\frac{1}{100}\sum_{i=1}^{99}i=99-\frac{1}{100}\cdot\frac{(99)(100)}{2}=49.5,$$

$$\sum_{i=1}^{99}p_iq_i=\frac{1}{100}\sum_{i=1}^{99}i-\frac{1}{(100)^2}\sum_{i=1}^{99}i^2$$

$$=49.5-\frac{1}{(100)^2}\cdot\frac{(99)(100)(199)}{6}=16.665.$$

由中心极限定理可知，回答正确的问题总数 $\sum_{i=1}^{99}X_i$ 近似地服从正态分布，其均值为49.5，方差为 $(16.665)^{1/2}=4.08$. 因此，随机变量

$$Z=\frac{\sum_{i=1}^{n}X_i-49.5}{4.08}$$

近似服从标准正态分布. 于是，

$$\Pr\left(\sum_{i=1}^{n}X_i\geqslant 60\right)=\Pr(Z\geqslant 2.5735)\approx 1-\Phi(2.5735)=0.0050.$$ ◀

5.7.4　扩展研究：中心定理证明的概要

设 $X_1, X_2, \cdots$ 是一列随机变量，对于 $n = 1, 2, \cdots$，以 F_n 表示 X_n 的分布函数，X^* 表示另一个随机变量，其分布函数为 F^*. 假定 F^* 是整个实轴上的连续函数. 则如果满足

$$\lim_{n\to\infty} F_n(x) = F^*(x) \quad 对于 -\infty < x < \infty. \tag{5.7.13}$$

则称随机变量序列 $X_1, X_2, \cdots$ 依分布收敛于随机变量 X^*.

有时，也简单地称 X_n 依分布收敛于 X^*，称 X^* 的分布为 X_n 的近似分布. 这样，根据林德伯格－莱维中心极限定理，如(5.7.1)式所表示的，随机变量 $n^{1/2}(\overline{X}_n - \mu)/\sigma$ 依分布收敛于服从标准正态分布的随机变量；或等价地，$n^{1/2}(\overline{X}_n - \mu)/\sigma$ 的近似分布为标准正态分布.

288

1. 矩母函数的收敛

由于下面的定理的原因，矩母函数在研究依分布收敛时是非常重要的，该定理的证明由于太难在此不予给出.

定理 5.7.4　*设 $X_1, X_2, \cdots$ 是随机变量序列，对于 $n = 1, 2, \cdots$，以 F_n 表示 X_n 的分布函数，ψ_n 表示 X_n 的矩母函数.*

以 X^* 表示另一个随机变量，其分布函数为 F^*，矩母函数为 ψ^*. 假定对于所有的 t，矩母函数 $\psi_n (n = 1, 2, \cdots)$ 和 ψ^* 都存在. 如果对于在 $t = 0$ 的某个区间内的所有 t，$\lim_{n\to\infty} \psi_n(t) = \psi^*(t)$，则 $X_1, X_2, \cdots$ 依分布收敛于 X^*. ■

换句话说，如果相应的矩母函数序列 $\psi_1, \psi_2, \cdots$ 收敛于矩母函数 ψ^*，则分布函数序列 $F_1, F_2, \cdots$ 必定收敛于分布函数 F^*.

2. 定理 5.7.1 证明的概要

现在，我们要给出林德伯格－莱维中心极限定理证明的概要. 设随机变量 $X_1, \cdots, X_n$ 为来自于均值为 μ、方差为 σ^2 的总体分布的一个样本. 为方便起见，假定该分布的矩母函数存在，尽管没有该假设，中心极限定理也成立.

对于 $i = 1, \cdots, n$，令 $Y_i = (X_i - \mu)/\sigma$，则随机变量 $Y_1, \cdots, Y_n$ 独立同分布，且均值都为 0 和方差都为 1. 并且，令

$$Z_n = \frac{n^{1/2}(\overline{X}_n - \mu)}{\sigma} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} Y_i.$$

下面通过证明 Z_n 的矩母函数收敛于标准正态分布的矩母函数，来证明如(5.7.1)式所表示的 Z_n 依分布收敛于服从标准正态分布的随机变量.

若以 $\psi(t)$ 表示每个随机变量 $Y_i (i = 1, \cdots, n)$ 的矩母函数，则由定理 4.4.3 知，$\sum_{i=1}^{n} Y_i$ 的矩母函数为 $[\psi(t)]^n$. 而且，由定理 4.4.2，Z_n 的矩母函数 $\zeta_n(t)$ 为

$$\zeta_n(t) = \left[\psi\left(\frac{t}{n^{1/2}}\right)\right]^n.$$

在这个问题中，$\psi'(0)=\mathrm{E}(Y_i)=0,\psi''(0)=\mathrm{E}(Y_i^2)=1$. 因此，$\psi(t)$在$t=0$处的泰勒级数展开式具有下述形式：

$$\psi(t)=\psi(0)+t\psi'(0)+\frac{t^2}{2!}\psi''(0)+\frac{t^3}{3!}\psi'''(0)+\cdots=1+\frac{t^2}{2}+\frac{t^3}{3!}\psi'''(0)+\cdots.$$

则

$$\zeta_n(t)=\left[1+\frac{t^2}{2n}+\frac{t^3\psi'''(0)}{3!\ n^{3/2}}+\cdots\right]^n.$$

289

由高等微积分可知，对于某些 a_n 和 b，如果 $\lim_{n\to\infty}a_n=b$，则

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n=\mathrm{e}^b.$$

但是

$$\lim_{n\to\infty}\left[\frac{t^2}{2}+\frac{t^3\psi'''(0)}{3!\ n^{1/2}}+\cdots\right]=\frac{t^2}{2}.$$

因此，

$$\lim_{n\to\infty}\zeta_n(t)=\exp\left(\frac{1}{2}t^2\right).\tag{5.7.14}$$

由于(5.7.14)式左边是标准正态分布的矩母函数，由定理5.7.4知，Z_n 的近似分布一定是标准正态分布.

李雅普洛夫中心极限定理证明的概要也可以用与上述相同的步骤来证明，但这里不深入进行讨论.

5.7.5 小结

本节给出中心极限定理的两种形式. 这两个定理的结论是：大量独立随机变量平均值的分布非常接近于正态分布. 一个定理要求所有的随机变量都有相同分布且具有有限方差. 另一个定理不要求随机变量同分布，但要求三阶矩存在，且满足(5.7.7)的条件. δ 方法给出样本均值的平滑函数近似分布的一种求法.

习题

1. 设一台机器每分钟生产绳子的长度服从均值为4英尺、标准差为5英寸的正态分布. 假定不同时间里生产绳子的长度是相互独立且同分布的. 试求该机器一小时内至少生产250英尺绳子的近似概率.
2. 设在一个大都市里有75%的人住在市区，25%的人住在郊区. 如果参加某个音乐会的1200人代表来自都市的一个随机样本. 求在参加音乐会的人当中，来自郊区的人数少于270的概率.
3. 设在任意给定的一卷布中，瑕疵数服从均值为5的泊松分布. 现随机抽取了125卷布构成一个随机样本，记录下每卷布中的瑕疵数. 试求每卷布平均的瑕疵数小于5.5的

概率.

4. 假定一个容量为 n 的随机样本来自于均值为 μ、标准差为3的分布总体. 利用中心极限定理来确定满足下列关系最小的 n:

290

$$\Pr(|X_n - \mu| < 0.3) \geqslant 0.95.$$

5. 设一个大型制造厂生产出有缺陷的产品比例为0.1. 现从该产品中随机抽取一样本，为使样本中有缺陷的产品的比例小于0.13的概率至少为0.99，则最小的样本容量为多少?

6. 设有三个女孩 A、B 和 C 朝同一目标扔雪球. 假定 A 扔10次，每次击中的概率为0.3; B 扔15次，每次击中的概率为0.2; C 扔20次，每次击中的概率为0.1. 计算目标至少被击中12次的概率.

7. 设从集合{0,…,9}中随机有放回地抽取16个数字，则这16个数字的平均数在4~6之间的概率为多少?

8. 设参加某宴会的人喝的饮料取自于一个包含63盎司液体的瓶子中. 假定每次所倒的饮料重量的期望值为2盎司，标准差为1/2盎司，设每一次倒的饮料重量是相互独立的. 计算倒了36次后饮料瓶不空的概率.

9. 一物理学家对某物体的比重进行了25次独立的测量. 该物理学家清楚装置的局限性，每次测量的标准差为 σ 个单位.

   a. 利用切比雪夫不等式，求他测量的平均值与物体真实比重的差小于 $\sigma/4$ 个单位的概率的下限.

   b. 利用中心极限定理，计算a.中概率的近似值.

10. 一容量为 n 的随机样本来自于均值为 μ、标准差为 σ 的分布总体.

    a. 利用切比雪夫不等式确定满足下列关系式最小的 n:

$$\Pr\left(|X_n - \mu| \leqslant \frac{\sigma}{4}\right) \geqslant 0.99.$$

    b. 利用中心极限定理确定最小的 n，使a.中的关系式近似成立.

11. 设在某所大学里毕业班中平均有1/3学生的父母都来参加毕业典礼，另有1/3学生的父母只有一人参加毕业典礼，剩下的1/3学生的父母均不来参加毕业典礼. 如果在某个班级里，有600个毕业学生，则有不多于650个父母参加毕业典礼的概率为多少?

12. 令 X_n 表示服从参数为 n 和 p_n 的二项分布的随机变量. 假定 $\lim_{n\to\infty} np_n = \lambda$. 证明: X_n 的矩母函数收敛于均值为 λ 的泊松分布的矩母函数.

13. 设 $X_1,\cdots,X_n$ 为来自均值为 θ、方差为 σ^2 的正态总体的一个样本，其中 θ 未知，且 $\theta \neq 0$. 试确定 $\overline{X}_n^3$ 的近似分布.

14. 设 $X_1,\cdots,X_n$ 为来自均值为0、方差为 σ^2 的正态总体的一个样本，其中 σ^2 未知.

    a. 求统计量 $\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^{-1}$ 的近似分布.

    b. 找出统计量 $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ 的一个方差稳定化变换.

5.8 对连续性的修正

中心极限定理的某些应用允许我们通过计算正态随机变量在区间$[a,b]$上的概率，来近似离散型随机变量 X 在$[a,b]$上的概率. 该近似可以通过小心地近似 $\Pr(X=a)$和 $\Pr(X=b)$的方法来得到稍微的改进.

291

5.8.1 用连续分布近似离散分布

设 $X_1,\cdots,X_n$ 为来自离散型分布的一个随机样本，令$X=X_1+\cdots+X_n$. 由前面几节可以看到，即使 X 的分布是离散的，也可以用正态分布这一连续型分布来近似 X 的分布. 在本节，将介绍一个标准的方法，该方法用于提高由连续型分布的概率近似离散型分布概率的精度.

为此，假定离散型随机变量 X 具有概率函数 $f(x)$，我们希望它能够用概率密度函数为 $g(x)$的连续型分布来近似. 为简单起见，只考虑 X 的所有可能取值为整数的离散型分布. 本章所介绍的二项分布、超几何分布、泊松分布和负二项分布均满足该条件.

如果概率密度函数 $g(x)$是 X 的分布很好的近似，则对于所有的整数 a, b, 概率

$$\Pr(a\leqslant X\leqslant b)=\sum_{x=a}^{b}f(x) \tag{5.8.1}$$

可以通过积分

$$\int_a^b g(x)\,\mathrm{d}x. \tag{5.8.2}$$

进行简单的近似. 事实上，例 5.7.1 和例 5.7.6 已经用过这个近似，其中的 $g(x)$就是由中心极限定理得到的近似于正态分布的概率密度函数.

该简单的近似有下述缺点：虽然对于离散型分布来说，$\Pr(X\geqslant a)$和 $\Pr(X>a)$有不同的值，但对于连续型分布，这两个概率值总是相等的. 该缺点的另一种表述就是：虽然对于 X 的每一个可能整数值 x，有 $\Pr(X=x)>0$，但对于近似正态分布的概率密度函数，该概率一定是零.

5.8.2 近似直方图

可以用直方图、条形图来表示随机变量 X 的概率函数 $f(x)$，如图 5-6 所描绘. 对于每个整数 x，用 $x-1/2$ 到 $x+1/2$ 为底、$f(x)$为高的矩形面积来表示 x 的概率. 如此一来，可以将底以 x 为中心的矩形面积简单表示成 $f(x)$，近似概率密度函数 $g(x)$的图形也如图 5-6 表示.

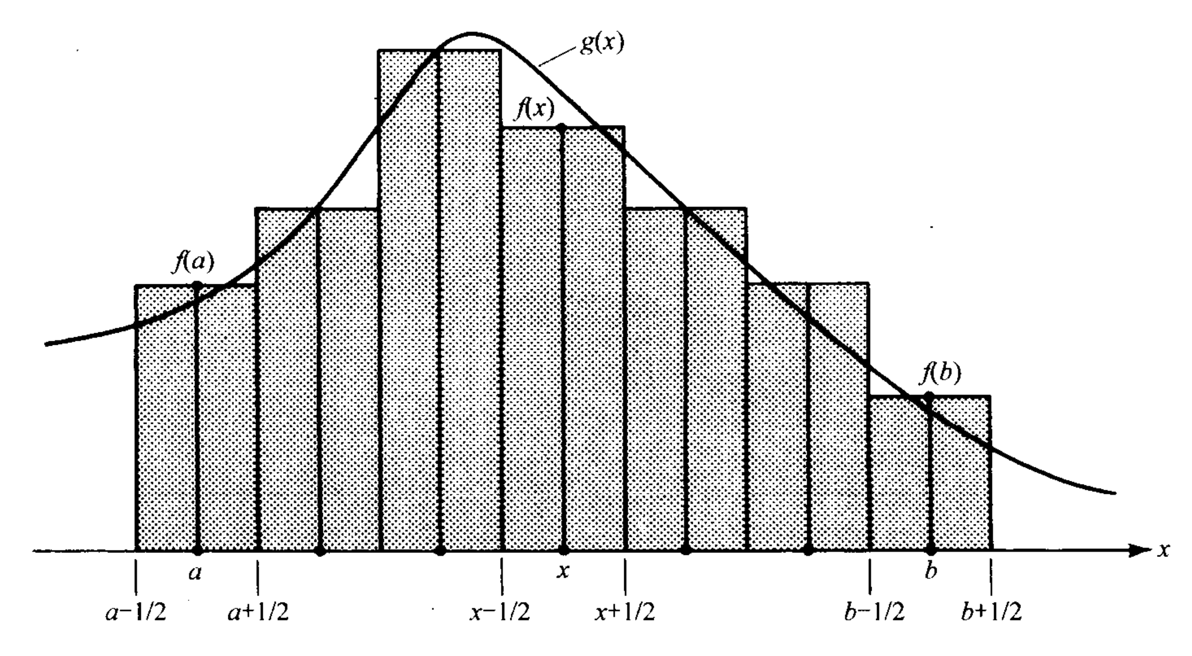


图 5-6　以概率密度函数近似直方图

根据该观点，可以看到如(5.8.1)式所表示的，$\Pr(a\leqslant x\leqslant b)$是在图 5-6 中底以 $a,a+1,\cdots,b$ 为中心的矩形的面积和. 由图 5-6 可以看到，这些面积的和由积分

292

$$\int_{a-(1/2)}^{a+(1/2)} g(x)\,\mathrm{d}x. \tag{5.8.3}$$

来近似. 由(5.8.2)式到(5.8.3)式的调整称作对于连续性的修正.

如果利用对连续性修正的方法，可以近似地找到在某个整数 a 处的概率 $f(a)$ 如下：

$$\Pr(X=a)=\Pr\left(a-\frac{1}{2}\leqslant x\leqslant a+\frac{1}{2}\right)\approx\int_{a-(1/2)}^{a+(1/2)} g(x)\,\mathrm{d}x. \tag{5.8.4}$$

类似地，

$$\Pr(X>a)=\Pr(X\geqslant a+1)=\Pr\left(X\geqslant a+\frac{1}{2}\right)\approx\int_{a+(1/2)}^{\infty} g(x)\,\mathrm{d}x. \tag{5.8.5}$$

例 5.8.1　考试问题　为了解释对于连续性修正方法的应用，现重新考虑例 5.7.6. 在那个例子中，一门考试包含 99 个不同难度的问题，要确定的是 $\Pr(X\geqslant 60)$，其中，X 表示某学生答对的题目总数. 在那个例子的假设下，根据中心极限定理，离散型随机变量 X 的分布可用均值为 49.5、标准差为 4.08 的正态分布来近似.

293

如果用对连续性修正的方法，可以得到：

$$\Pr(X\geqslant 60)=\Pr(X\geqslant 59.5)=\Pr\left(Z\geqslant\frac{59.5-49.5}{4.08}\right)\approx 1-\Phi(2.4510)=0.007.$$

而在 5.7 节中没有任何修正得到的值为 0.005. ◀

例 5.8.2　抛掷硬币　设抛掷一枚均匀硬币 20 次，假定所有的抛掷是相互独立的. 则恰好出现 10 次正面的概率为多少？

以 X 表示在 20 次投掷中出现正面的总次数. 根据中心极限定理，X 近似地服从正态分布，其均值为 10，标准差为 $[(20)(1/2)(1/2)]^{1/2}=2.236$. 如果利用对于连续性的修正，

$$\Pr(X=10)=\Pr(9.5\leqslant X\leqslant 10.5)=\Pr\left(-\frac{0.5}{2.236}\leqslant Z\leqslant\frac{0.5}{2.236}\right)$$
$$\approx\Phi(0.2236)-\Phi(-0.2236)=0.177.$$

根据本书后面的附录给出的二项分布表，得到 $\Pr(X=10)$ 的精确值为 0.1762. 这样，用对于连续性修正的正态近似方法是非常好的. ◀

5.8.3 小结

以 X 表示只取整数值的随机变量. 设 X 近似地服从均值为 μ、方差为 σ^2 的正态分布. 令 a 和 b 是整数，希望近似得到概率 $\Pr(a\leqslant X\leqslant b)$. 利用对正态分布的连续性修正，该近似值为 $\Phi([b+1/2-\mu]/\sigma)-\Phi([a-1/2-\mu]/\sigma)$，而不是 $\Phi([b-\mu]/\sigma)-\Phi([a-\mu]/\sigma)$.

习题

1. 令 $X_1,\cdots,X_{30}$ 是相互独立的随机变量，且每个都有离散型分布，概率函数为：

$$f(x)=\begin{cases}1/4 & \text{如果 } x=0 \text{ 或 } 2,\\ 1/2 & \text{如果 } x=1,\\ 0 & \text{其他.}\end{cases}$$

   利用中心极限定理和连续性修正来近似 $X_1+\cdots+X_{30}$ 至多为 33 的概率.
2. 令 X 表示 15 重伯努利试验中成功的次数，在每次试验中成功概率为 $p=0.3$.
   a. 利用连续性修正的中心极限定理，求 $\Pr(X=4)$ 的近似值.
   b. 将该概率的精确值与 a. 计算得到的答案进行比较.

294

3. 利用连续性修正，计算例 5.7.1 中的概率.
4. 利用连续性修正，计算 5.7 节习题 2 中的概率.
5. 利用连续性修正，计算 5.7 节习题 3 中的概率.
6. 利用连续性修正，计算 5.7 节习题 6 中的概率.
7. 利用连续性修正，计算 5.7 节习题 7 中的概率.

5.9 伽玛分布

伽玛分布族是随机变量取值为正的一种常见模型. 指数分布族是伽玛分布族中的一类特殊分布族. 泊松过程中，相邻事件发生的时间间隔具有指数分布. 伽玛函数与伽玛分布相联系，它将整数的阶乘推广到所有正数.

5.9.1 伽玛函数

对于每个正数 α，令 $\Gamma(\alpha)$ 的值由下述积分所定义：

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx. \tag{5.9.1}$$

可以看到，对于每个 $\alpha>0$ 的值，该积分的值都是有限的. 对于 $\alpha>0$，称由(5.9.1)式定义的函数 Γ 为伽玛函数. 下面，我们给出伽玛函数的一些性质.

定理 5.9.1 如果 $\alpha>1$，则

$$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1). \tag{5.9.2}$$

证明： 我们用分步积分的方法证明(5.9.1)式. 如果令 $u=x^{\alpha-1}$，$dv=e^{-x}dx$，则 $du=(\alpha-1)x^{\alpha-2}dx, v=-e^{-x}$. 因此，

$$\begin{aligned}\Gamma(\alpha)&=\int_0^\infty u\,dv=[uv]_0^\infty-\int_0^\infty v\,du\\&=[-x^{\alpha-1}e^{-x}]_0^\infty+(\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}\,dx\\&=0+(\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

295

■

由定理5.9.1，对于每个 $n\geqslant 2$ 的整数，

$$\begin{aligned}\Gamma(n)&=(n-1)\Gamma(n-1)=(n-1)(n-2)\Gamma(n-2)\\&=(n-1)(n-2)\cdots 1\cdot\Gamma(1)=(n-1)!\ \Gamma(1).\end{aligned}$$

并且，由(5.9.1)式，

$$\Gamma(1)=\int_0^\infty e^{-x}\,dx=1.$$

因此，对于 $n=2,3,\cdots$，$\Gamma(n)=(n-1)!$，且 $\Gamma(1)=1=0!$，由此，我们可得到如下结论：

定理 5.9.2 对于每个正整数 n，

$$\Gamma(n)=(n-1)!. \tag{5.9.3}$$

■

在许多统计应用中，当 α 是正整数或者具有形式 $\alpha=n+(1/2)$（n 为正整数）时，需要计算 $\Gamma(\alpha)$. 由(5.9.2)式知，对于每个正整数 n，有

$$\Gamma\left(n+\frac{1}{2}\right)=\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right). \tag{5.9.4}$$

因此，如果能确定 $\Gamma\left(\frac{1}{2}\right)$ 的值，就能确定 $\Gamma\left(n+\frac{1}{2}\right)$ 的值.

由(5.9.1)式，

$$\Gamma\left(\frac{1}{2}\right)=\int_0^\infty x^{-1/2}e^{-x}\,dx.$$

如果在上述积分中令 $x=(1/2)y^2$，则 $dx=y\,dy$，且

$$\Gamma\left(\frac{1}{2}\right)=2^{1/2}\int_0^\infty \exp\left(-\frac{1}{2}y^2\right)dy. \tag{5.9.5}$$

由于标准正态分布的概率密度函数在整个定义区间上的积分等于1，因此，有

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y^2\right)dy = (2\pi)^{1/2}.$$

于是，

$$\int_{0}^{\infty} \exp\left(-\frac{1}{2}y^2\right)dy = \frac{1}{2}(2\pi)^{1/2} = \left(\frac{\pi}{2}\right)^{1/2}.$$

由(5.9.5)式，

$$\Gamma\left(\frac{1}{2}\right) = \pi^{1/2}. \tag{5.9.6}$$

比如说，由(5.9.4)和(5.9.6)式，

296

$$\Gamma\left(\frac{7}{2}\right) = \left(\frac{5}{2}\right)\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\pi^{1/2} = \frac{15}{8}\pi^{1/2}.$$

5.9.2 伽玛分布

对于 $\alpha>0$，$\beta>0$，如果连续型随机变量 X 的概率密度函数具有下述形式：

$$f(x\mid\alpha,\beta) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & \text{对于 } x>0, \\ 0 & \text{对于 } x\leqslant 0. \end{cases} \tag{5.9.7}$$

则称 X 服从参数为 α 和 β 的伽玛分布. 此概率密度在整个区间上的积分等于1，这是因为根据伽玛函数的定义，

$$\int_0^\infty x^{\alpha-1}e^{-\beta x}dx = \frac{\Gamma(\alpha)}{\beta^\alpha}. \tag{5.9.8}$$

如果 X 服从参数为 α 和 β 的伽玛分布，则由(5.9.7)式和(5.9.8)式可以得到 X 的矩. 对于 $k=1,2,\cdots$，有

$$\begin{aligned} E(X^k) &= \int_0^\infty x^k f(x\mid\alpha,\beta)\,dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+k-1}e^{-\beta x}dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}} = \frac{\Gamma(\alpha+k)}{\beta^k\Gamma(\alpha)} \\ &= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{\beta^k}. \end{aligned}$$

因此，特别地，有

$$E(X) = \frac{\alpha}{\beta}$$

和

$$\mathrm{Var}(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}.$$

用同样的方法可以得到 X 的矩母函数 ψ，具体如下：

$$\psi(t)=\int_0^{\infty}\mathrm{e}^{tx}f(x\mid\alpha,\beta)\,\mathrm{d}x=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{\alpha-1}\mathrm{e}^{-(\beta-t)x}\mathrm{d}x.$$

对于每个满足 $t<\beta$ 的 t，上述积分都是有限的. 因此，由(5.9.8)式，可以得到

$$\psi(t)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha)}{(\beta-t)^{\alpha}}=\left(\frac{\beta}{\beta-t}\right)^{\alpha}\quad 对于\ t<\beta.\tag{5.9.9}$$

297

现在来证明具有相同参数 β 的伽玛分布且独立的随机变量的和也具有伽玛分布.

定理5.9.3 如果随机变量 $X_1,\cdots,X_k$ 相互独立，且 X_i 服从参数为 α_i 和 $\beta(i=1,\cdots,k)$ 的伽玛分布，则 $X_1+\cdots+X_k$ 服从参数为 $\alpha_1+\cdots+\alpha_k$ 和 β 的伽玛分布.

证明： 如果以 ψ_i 表示 X_i 的矩母函数，则由(5.9.9)式，对于 $i=1,\cdots,k$，

$$\psi_i(t)=\left(\frac{\beta}{\beta-t}\right)^{\alpha_i}\quad 对于\ t<\beta.$$

若以 ψ 表示 $X_1+\cdots+X_k$ 的矩母函数，则由定理4.4.3，得

$$\psi(t)=\prod_{i=1}^{k}\psi_i(t)=\left(\frac{\beta}{\beta-t}\right)^{\alpha_1+\cdots+\alpha_k}\quad 对于\ t<\beta.$$

可以将矩母函数 ψ 看成是参数为 $\alpha_1+\cdots+\alpha_k$ 和 β 的伽玛分布的矩母函数. 因此，$X_1+\cdots+X_k$ 也服从伽玛分布. ■

5.9.3 指数分布

对于 $\beta>0$，如果连续型随机变量 X 的概率密度函数 $f(x\mid\beta)$ 具有下述形式：

$$f(x\mid\beta)=\begin{cases}\beta\mathrm{e}^{-\beta x} & 对于\ x>0,\\ 0 & 对于\ x\leqslant 0.\end{cases}\tag{5.9.10}$$

则称 X 服从参数为 β 的指数分布. 由(5.9.10)式可以看出，具有参数为 β 的指数分布与参数为 $\alpha=1$ 和 β 的伽玛分布是一样的. 因此，如果随机变量服从参数为 β 的指数分布，则由伽玛分布可以得到

$$\mathrm{E}(X)=\frac{1}{\beta},\quad \mathrm{Var}(X)=\frac{1}{\beta^2}.$$

类似地，由(5.9.9)式可以得到 X 的矩母函数为：

$$\psi(t)=\frac{\beta}{\beta-t}\quad 对于\ t<\beta.$$

5.9.4 指数分布的无记忆性

指数分布具有类似于5.5节中所介绍的几何分布的无记忆性. 可以将此性质表述为：如果 X 服从参数为 β 的指数分布，则对于 $t>0$，有

298

$$\Pr(X\geqslant t)=\int_t^{\infty}\beta\mathrm{e}^{-\beta x}\,\mathrm{d}x=\mathrm{e}^{-\beta t}.\tag{5.9.11}$$

因此，对于 $t>0$，$h>0$，

$$\Pr(X \geqslant t+h \mid X \geqslant t) = \frac{\Pr(X \geqslant t+h)}{\Pr(X \geqslant t)} = \frac{e^{-\beta(t+h)}}{e^{-\beta t}} = e^{-\beta h} = \Pr(X \geqslant h). \quad (5.9.12)$$

可以证明（见习题23）指数分布是连续型分布中唯一的具有无记忆性的分布.

为了解释指数分布的无记忆性，可以假定 X 表示某个事件发生前所经历的分钟数. 根据(5.9.12)式，如果事件在 t 分钟内还未发生，则事件在接下来的 h 分钟内仍未发生的概率为 $e^{-\beta h}$. 这与从0时刻开始，事件在 h 分钟内未发生的概率是相同的. 即，不考虑事件未发生所经历的时间，事件在紧接着的 h 分钟内发生的概率总是具有相同的值. 因此，从理论上讲，为了计算事件在未来发生的概率，没有必要考虑事件过去所发生的情况.

严格地讲，并不是所有的实际问题都满足无记忆性. 比如说，设随机变量 X 表示一个电灯泡在烧坏之前所使用的时间，电灯泡在将来期望使用的时间长度依赖于过去已经使用的时间长度. 然而，指数分布作为表示产品寿命随机变量的近似分布却是很有效的.

5.9.5 寿命测试

假设为了测定 n 个电灯泡的寿命，让它们同时开始使用. 设 n 个电灯泡烧坏是相互独立的，且每个电灯泡的寿命都服从参数为 β 的指数分布. 即，对于 $i=1,\cdots,n$，如果以 X_i 表示第 i 个灯泡的使用寿命，则随机变量 $X_1,\cdots,X_n$ 独立同分布，且都服从参数为 β 的指数分布. 现在来确定直到 n 个电灯泡中第一个烧坏所经历的时间 Y_1 的分布.

由于第一个灯泡烧坏的时间 Y_1 等于 n 个寿命 $X_1,\cdots,X_n$ 中最小的一个，所以，可以写成 $Y_1=\min\{X_1,\cdots,X_n\}$. 对于 $t>0$，

$$\begin{aligned}\Pr(Y_1>t) &= \Pr(X_1>t,\cdots,X_n>t)\\ &= \Pr(X_1>t)\cdots\Pr(X_n>t)\\ &= e^{-\beta t}\cdots e^{-\beta t} = e^{-n\beta t}.\end{aligned}$$

将该结果与(5.9.11)式比较，可以看到 Y_1 一定服从参数为 $n\beta$ 的指数分布. 299

概括起来，可以得到如下结果：

定理5.9.4 *设变量 X_1，…，X_n 为来自于参数为 β 的指数分布总体一个随机样本，则 $Y_1=\min\{X_1,\cdots,X_n\}$服从参数为 $n\beta$ 的指数分布.* ■

下面，来确定第一个电灯泡烧坏和第二个电灯泡烧坏的时间间隔 Y_2 的分布.

一个电灯泡烧坏后，其余 $n-1$ 个电灯泡仍在使用. 而且，不考虑第一个电灯泡烧坏的时间，由指数分布的无记忆性，其余 $n-1$ 个电灯泡中每一个的剩余寿命仍服从参数为 β 的指数分布. 即与从 $t=0$ 重新开始对 $n-1$ 个新灯泡进行测试是相同的. 因此，Y_2 等于 $n-1$ 个独立同分布的随机变量中最小的一个，其中每一个随机变量都服从参数为 β 的指数分布. 由定理5.9.4，Y_2 服从参数为$(n-1)\beta$ 的指数分布.

按此方法继续下去，可以得到第二个电灯泡烧坏和第三个电灯泡烧坏的时间间

隔 Y_3 服从参数为 $(n-2)\beta$ 的指数分布. 最后，只剩下一个电灯泡还未烧坏，直到它烧坏的时间间隔服从参数为 β 的指数分布.

5.9.6 与泊松过程的关系

指数分布通常用在实际问题中表示某事件发生前所经历时间的分布. 比如说，用该分布表示一台机器或电子元件无故障运行的时间，在某服务台一个顾客接受服务的时间，及在某服务中两个顾客连续到达的时间间隔.

如果考虑的事件发生是按照泊松过程进行的，正如5.4节所定义的，则直到第一个事件发生的等待时间及两个相邻发生事件的时间间隔都服从指数分布. 该事实对于解释指数分布为什么可以用在许多实际问题中提供了有利的理论支持.

例5.9.1 放射性颗粒 考虑例5.4.3泊松过程的例子，假定，我们感兴趣的是直到第一个放射性颗粒击中目标所要等待的时间. 设 Y_1 表示直到第一个颗粒击中目标经历的时间，X 表示在长度为 t 的时间间隔内，击中目标的颗粒数. 容易看到，$Y_1 \leqslant t$ 当且仅当 $X \geqslant 1$. 即，到时间 t 为止，第一个颗粒击中目标当且仅当至少有一个颗粒击中目标. 我们已经知道，X 服从均值为 λt 的泊松分布，其中的 λ 为过程的速率. 因此，对于 $t>0$，

300

$$\Pr(Y_1 \leqslant t) = \Pr(X \geqslant 1) = 1 - \Pr(X=0) = 1 - e^{-\lambda t}.$$

即，Y_1 的分布函数为:

$$F(y) = \begin{cases} 1 - e^{-\lambda t} & \text{对于 } t>0, \\ 0 & \text{对于 } t \leqslant 0. \end{cases}$$

对该分布函数求导，得到 Y_1 的概率密度函数:

$$f(y) = \begin{cases} \lambda e^{-\lambda t} & \text{对于 } t>0, \\ 0 & \text{对于 } t \leqslant 0. \end{cases}$$

很容易看到，这就是参数为 λ 的指数分布. ◀

注: 泊松过程中事件发生的时间间隔 显然，例5.9.1中的结论不依赖于放射性颗粒任何特殊的性质. 事实上，在速率为 λ 的泊松过程中，直到第一个事件发生的时间的分布都相同，都服从参数为 λ 的指数分布. 类似地但稍微复杂一点的结论可以证明: 相邻事件发生的时间间隔是相互独立的，且都服从参数为 λ 的指数分布. 将该结论与定理5.9.3结合起来，可以看到，指数分布是伽玛分布的事实表明: 在速率为 λ 的泊松过程中直到第 k 个事件发生的等待时间服从参数为 k 和 λ 的伽玛分布.

5.9.7 小结

定义伽玛函数为: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$，它具有性质: 对于 $n=1,2,\cdots$，$\Gamma(n) = (n-1)!$.

如果随机变量 $X_1,\cdots,X_n$ 相互独立，且具有相同的第二个参数 β，则 $\sum_{i=1}^{n} X_i$ 也服从伽玛分布，其中第一参数为 $X_1,\cdots,X_n$ 的第一个参数的和，第二个参数为 β. 具有参数为 β 的指数分布与参数为 1 和 β 的伽玛分布是一样的. 因此，来自于参数为 β 的指数分布总体的一个样本中的 n 个随机变量的和服从参数为 n 和 β 的伽玛分布. 对于具有速率 λ 的泊松过程，相邻事件发生的时间间隔服从参数为 λ 的指数分布，且相互独立. 直到第 k 个事件发生的等待时间服从参数为 k 和 λ 的伽玛分布.

习题

1. 设随机变量 X 服从参数为 α 和 β 的伽玛分布，c 是一个正的常数. 证明：cX 服从参数为 α 和 β/c 的伽玛分布.
2. 计算具有参数为 β 的指数分布的分位数函数.
3. 画出具有下列参数 α 和 β 的伽玛分布的概率密度曲线：a. $\alpha=1/2$，$\beta=1$，b. $\alpha=1$，$\beta=1$，c. $\alpha=2$，$\beta=1$.
4. 计算具有参数为 α 和 β 的伽玛分布的众数.
5. 画出具有下列参数 β 的指数分布的概率密度曲线：a. $\beta=1/2$，b. $\beta=1$，c. $\beta=2$. [301]
6. 设随机变量 X_1，…，X_n 是来自参数为 β 的指数分布总体的一个随机样本. 求样本均值 $\overline{X}_n$ 的分布.
7. 假定一名顾客在超市柜台等候服务的时间(单位:分钟)服从均值为 3 的指数分布. 利用中心极限定理，试求在一个含有 16 名顾客的随机样本中，等候服务的总时间超过一个小时的概率.
8. 设随机变量 $X_1,\cdots,X_k$ 相互独立，且 X_i 服从参数为 $\beta_i(i=1,\cdots,k)$ 的指数分布. 令 $Y=\min\{X_1,\cdots,X_k\}$，证明：Y 服从参数为 $\beta_1+\cdots+\beta_k$ 的指数分布.
9. 设某系统中包含 3 个元件，每个元件都独立工作，且串联在一起，正如 3.7 节所描述的，只要系统中的一个元件毁坏，整个系统就崩溃. 假定第一个元件的寿命(单位:小时)服从参数为 $\beta=0.001$ 的指数分布；第二个元件的寿命服从参数为 $\beta=0.003$ 的指数分布；第三个元件的寿命服从参数为 $\beta=0.006$ 的指数分布. 求整个系统在 100 小时后崩溃的概率.
10. 设在一个含有 n 个相同电子元件的系统里，每个元件独立工作，且串联在一起，即只要系统中的一个元件毁坏，整个系统就崩溃. 假定每个元件的寿命(单位:小时)服从均值为 μ 的指数分布. 求直到整个系统崩溃所经历时间的均值和方差.
11. 设同时检验 n 件产品，每件产品是相互独立的，且每件产品的寿命均服从参数为 β 的指数分布. 求直到有 3 件产品毁坏所经历时间的期望值. 提示：所求的值为 $E(Y_1+Y_2+Y_3)$.
12. 重新考虑习题 10 中所描述的电子系统，但现在假定整个系统工作一直持续到有两个元件毁坏. 计算整个系统崩溃所经历时间的均值和方差.
13. 设有 5 名学生参加某门考试，每名学生相互独立，且每名学生完成考试的时间(单位:分钟)服从均值为 80 的指数分布. 假定考试在上午 9:00 开始. 求至少有一名学生在上午 9:40前完成考试的概率.
14. 重新考虑习题 13，5 名学生参加考试，设第一名学生在上午 9:25 完成考试. 求在上午 10:00前至少另有一名学生完成考试的概率.

15. 再一次考虑习题13，5名学生参加考试. 求在10分钟的时间间隔内，没有两名学生完成考试的概率.
16. 对于 $x_0>0$，$\alpha>0$，如果连续型随机变量 X 的概率密度函数 $f(x\mid x_0,\alpha)$ 具有下述形式：

$$f(x\mid x_0,\alpha)=\begin{cases}\dfrac{\alpha x_0^{\alpha}}{x^{\alpha+1}} & \text{对于 } x\geqslant x_0,\\ 0 & \text{对于 } x<x_0.\end{cases}$$

则称 X 服从参数为 x_0 和 α 的帕雷托（Pareto）分布. 证明：如果 X 服从参数为 x_0 和 α 的帕雷托分布，则随机变量 $\ln(X/x_0)$ 服从参数为 α 的指数分布.
17. 设随机变量 X 服从均值为 μ、方差为 σ^2 的正态分布. 求 $\mathrm{E}[(X-\mu)^{2n}]$，$n=1,2,\cdots$.
18. 考虑满足 $\Pr(X>0)=1$ 的随机变量 X，其概率密度函数为 f，分布函数为 F. 定义函数 h 如下：

$$h(x)=\frac{f(x)}{1-F(x)}\quad \text{对于 } x>0.$$

称函数 h 为 X 的失效率（failure rate）或疲劳函数（hazard function）. 证明：如果 X 服从指数分布，则对于 $x>0$，失效率 $h(x)$ 为常数.
19. 对于 $a>0$，$b>0$，如果连续型随机变量 X 的概率密度函数 $f(x\mid a,b)$ 具有下述形式：

$$f(x\mid a,b)=\begin{cases}\dfrac{b}{a^b}x^{b-1}\mathrm{e}^{-(x/a)^b} & \text{对于 } x>0,\\ 0 & \text{对于 } x\leqslant 0.\end{cases}$$

则称 X 服从参数为 a 和 b 的威布尔分布. 如果 X 具有这样的分布，则随机变量 X^b 服从参数为 $\beta=a^{-b}$ 的指数分布.

302

20. 如果习题18中定义的失效率 $h(x)$ 是一个关于 $x(x>0)$ 的递增函数，则称随机变量 X 具有上升的失效率. 如果 $h(x)$ 是一个关于 $x(x>0)$ 的递减函数，则称随机变量 X 具有下降的失效率. 设随机变量 X 如习题19所定义的，服从参数为 a 和 b 的威布尔分布. 证明：如果 $b>1$，X 具有上升的失效率；如果 $b<1$，X 具有下降的失效率.
21. 如果随机变量 X 服从参数为 $\alpha>2$ 和 $\beta>0$ 的伽玛分布.
    a. 证明：$1/X$ 的均值为 $\beta/(\alpha-1)$.
    b. 证明：$1/X$ 的方差为 $\beta^2/[(\alpha-1)^2(\alpha-2)]$.
22. 考虑例5.9.1中放射性粒子射击目标的泊松过程. 假定泊松过程的速率 λ 未知，但它服从参数为 α 和 β 的伽玛分布. 以 X 表示在 t 个时间单位里，击中目标的粒子数. 证明：在给定 $X=x$ 的条件下，λ 的条件分布为伽玛分布，并求伽玛分布的参数.
23. 设 F 为连续型分布函数，满足 $F(0)=0$，且假定具有分布函数 F 的分布具有无记忆性. 定义：$l(x)=\ln[1-F(x)]$.
    a. 证明：对于所有的 t，$h>0$，有

$$1-F(h)=\frac{1-F(t+h)}{1-F(t)}.$$

    b. 证明：对于所有的 t，$h>0$，$l(t+h)=l(t)+l(h)$.
    c. 证明：对于 $t>0$ 和所有的正整数 $k,m,l(kt/m)=(k/m)l(t)$.
    d. 证明：对于所有的 $t,c>0,l(ct)=cl(t)$.

e. 证明：对于 $t>0, g(t)=l(t)/t$ 是常数.

f. 证明：F 的分布函数一定是指数分布.

24. 回顾(5.6.12)式的Black-Scholes公式的推导过程. 本题中，假定股票在将来某个时刻 u 的价格为 $S_0e^{\mu u+W_u}$，其中的 W_u 服从参数为 αu 和 $\beta(\beta>1)$ 的伽玛分布. 以 r 表示无风险利率.

a. 证明：$e^{-ru}E(S_u)=S_0$ 当且仅当 $\mu=r-\alpha\ln(\beta/[\beta-1])$.

b. 设 $\mu=r-\alpha\ln(\beta/[\beta-1])$. 以 R 表示1减去参数为 αu 和1的伽玛分布函数. 证明：若在时刻 u 购买一股价格为 q 的这种股票，则该期权的风险中性价格为 $S_0R(c[\beta-1])-qe^{-ru}R(c\beta)$. 其中，

$$c=\ln\left(\frac{q}{S_0}\right)+\alpha u\ \ln\left(\frac{\beta}{\beta-1}\right)-ru.$$

c. 设 $u=1$，$q=S_0$，$r=0.06$，$\alpha=1$，$\beta=10$ 时，计算该期权的价格.

5.10 二元正态分布

将介绍的第一个多元连续型分布是正态分布在二维空间的推广. 二元正态分布比一元正态边际分布具有更复杂的结构.

5.10.1 二元正态分布的定义

设 Z_1 和 Z_2 是相互独立的随机变量，且均服从标准正态分布. 则对于它们所有可能的取值 z_1 和 z_2，Z_1 和 Z_2 的联合概率密度函数 $g(z_1,z_2)$ 具有下述形式：

$$g(z_1,z_2)=\frac{1}{2\pi}\exp\left[-\frac{1}{2}(z_1^2+z_2^2)\right]. \tag{5.10.1}$$

对于常数 $\mu_1,\mu_2,\sigma_1,\sigma_2$ 和 ρ，其中 $-\infty<\mu_i<\infty\ (i=1,2), \sigma_i>0\ (i=1,2), -1<\rho<1$，定义两个新的随机变量 X_1 和 X_2 如下：

$$X_1=\sigma_1Z_1+\mu_1,\quad X_2=\sigma_2[\rho Z_1+(1-\rho^2)^{1/2}Z_2]+\mu_2. \tag{5.10.2}$$

下面来推导 X_1 和 X_2 的联合概率密度函数 $f(x_1,x_2)$.

从 Z_1 和 Z_2 到 X_1 和 X_2 的变换是一个线性变换，Z_1 和 Z_2 系数矩阵的行列式 Δ 为 $\Delta=(1-\rho^2)^{1/2}\sigma_1\sigma_2$. 因此，如3.9节所讨论的，从 X_1 和 X_2 到 Z_1 和 Z_2 逆变换的雅可比行列式 J 为

$$J=\frac{1}{\Delta}=\frac{1}{(1-\rho^2)^{1/2}\sigma_1\sigma_2}. \tag{5.10.3}$$

由于 $J>0$，所以，$|J|$ 的值等于 J 本身. 由(5.10.2)式，如果用 X_1 和 X_2 表示 Z_1 和 Z_2 的解，则通过以 x_1 和 x_2 的表达式表示的 z_1 和 z_2 代替(5.10.1)式中的 z_1 和 z_2，然后乘上 $|J|$，可以得到联合概率密度函数 $f(x_1,x_2)$. 对于 $-\infty<x_1<\infty$，$-\infty<x_2<\infty$，可以证得如下结果：

$$f(x_1,x_2)=\frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}. \tag{5.10.4}$$

当随机变量 X_1 和 X_2 的联合概率密度函数具有(5.10.4)式的形式时，称随机变量 X_1 和 X_2 服从二元正态分布. 很容易由(5.10.4)式和(5.10.2)式得到二元正态分布的均值和方差. 因为 Z_1 和 Z_2 相互独立，且均值都为0，方差都为1，因此，$E(X_1)=\mu_1, E(X_2)=\mu_2, Var(X_1)=\sigma_1^2$，$Var(X_2)=\sigma_2^2$. 而且，由(5.10.2)式，可以得到 $Cov(X_1,X_2)=\rho\sigma_1\sigma_2$. 则 X_1 和 X_2 的相关系数仅仅是 ρ. 概括起来，如果 X_1 和 X_2 服从二元正态分布，其概率密度函数由(5.10.4)式确定，则

$$E(X_i)=\mu_i,\quad Var(X_i)=\sigma_i^2,\quad i=1,2.$$

而且，

$$\rho(X_1,X_2)=\rho.$$

可以看到，把二元正态分布看成是由独立的服从标准正态分布的随机变量进行一定线性组合后得到分布是很方便的. 然而，应该强调的是，二元正态分布很直接、自然地产生于许多实际问题中. 比如说，许多人群的两个物理特征比如说身高和体重的联合分布可认为近似地服从二元正态分布；其他的人群中，每个人的两门相关考试成绩近似地服从二元正态分布.

例5.10.1 跳甲的测量 Lubischew(1962)报告了对于不同种类跳甲的物理特征的测量值. 其研究主要关注：是否可以用容易得到的测量的某些组合来区分不同种类. 对来自于一类毛克虫甲克虫(Chaetocnema heikertingeri)的31个个体组成的样本，测量每只跳甲第一块踝骨的第一个关节和第一块踝骨的第二个关节，得到图5-7的所描绘的散点图. 该图中包含三个椭圆，它们对应一个拟合出的二元正态分布. 三个椭圆分别包含拟合二元正态分布25%、50%和75%的概率，拟合分布的相关系数为0.64. ◀

5.10.2 边际分布和条件分布

1. 边际分布

我们继续假定随机变量 X_1 和 X_2 服从二元正态分布，其联合概率密度函数具有(5.10.4)式的形式. 在研究该分布的性质时，以(5.10.2)式的形式表示 X_1 和 X_2 是比较方便的，其中的 Z_1 和 Z_2 是独立的标准正态随机变量. 特别地，由于 X_1 和 X_2 都是 Z_1 和 Z_2 的线性组合，则根据该表示法及推论5.6.1，可以得到 X_1 和 X_2 的边际分布也是正态分布. 这样，对于 $i=1,2$，X_i 的边际分布是均值为 μ_i、方差为 σ_i^2 的正态分布.

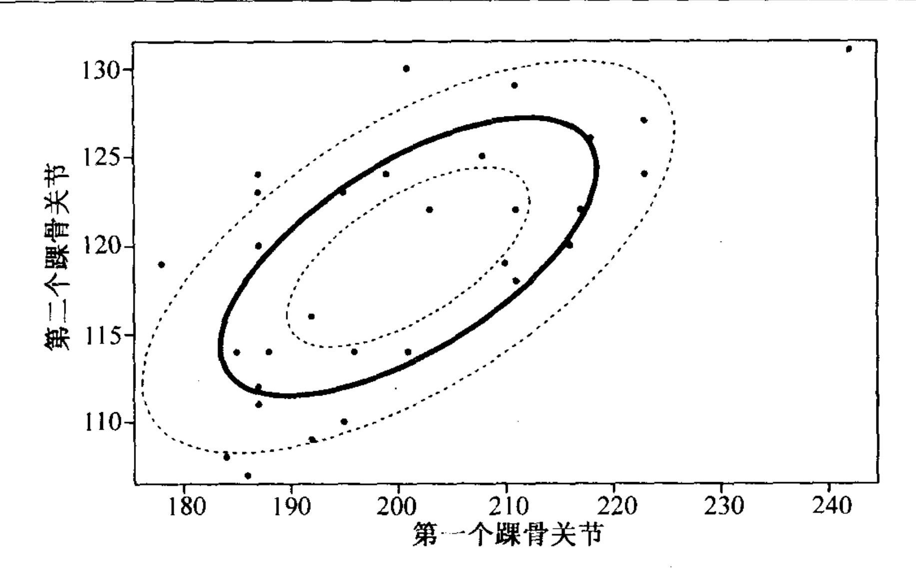


图 5-7 例 5.10.1 中以二元正态 25%，50% 和 75% 的概率得到的跳甲数据的散点图

2. 独立性和相关性

如果随机变量 X_1 和 X_2 不相关，则 $\rho=0$. 该种情形下，由(5.10.4)式，联合概率密度函数 $f(x_1,x_2)$ 可以分解成 X_1 的边际概率密度和 X_2 的边际概率密度的乘积. 因此，X_1 和 X_2 是独立的，且有下述结论： 314

具有二元正态分布的随机变量 X_1 和 X_2 相互独立当且仅当 X_1 和 X_2 不相关.

在 4.6 节中，已经看到，具有任意联合分布的随机变量 X_1 和 X_2 不独立也可以不相关.

3. 条件分布

由(5.10.2)式可以得到在给定 $X_1=x_1$ 的条件下 X_2 的条件分布. 如果 $X_1=x_1$，则 $Z_1=(x_1-\mu_1)/\sigma_1$. 因此，给定 $X_1=x_1$ 的条件下，X_2 的条件分布与下述随机变量的条件分布相同：

$$(1-\rho^2)^{1/2}\sigma_2 Z_2+\mu_2+\rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right) \tag{5.10.5}$$

因为 Z_2 服从标准正态分布，且与 X_1 独立，则由(5.10.5)式，在给定 $X_1=x_1$ 的条件下，X_2 的条件分布是正态分布，其均值为

$$\mathrm{E}(X_2\mid X_1)=\mu_2+\rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right), \tag{5.10.6}$$

方差为 $(1-\rho^2)\sigma_2^2$.

不能容易地由(5.10.2)式得到给定 $X_2=x_2$ 的条件下 X_1 的条件分布，这是因为在(5.10.2)式中，Z_1 和 Z_2 的形式不同. 然而，由(5.10.4)式可以看到，联合概率密度函数 $f(x_1,x_2)$ 对于两个变量 $(x_1-\mu_1)/\sigma_1$ 和 $(x_2-\mu_2)/\sigma_2$ 来说是对称的，因此，在求给定 $X_2=x_2$ 的条件下 X_1 的条件分布时，可以对给定 $X_1=x_1$ 的条件下 X_2 的条件分布（该分布刚得出）简单地交换 x_1 和 x_2、μ_1 和 μ_2、σ_1 和 σ_2 的位置得到. 如此一来， 315

在给定 $X_2=x_2$ 的条件下，X_1 的条件分布一定是正态分布，其均值为：

$$\mathrm{E}(X_1 \mid X_2)=\mu_1+\rho\sigma_1\left(\frac{x_2-\mu_2}{\sigma_2}\right), \tag{5.10.7}$$

方差为 $(1-\rho^2)\sigma_1^2$.

我们已经证明，二元正态分布的每个条件分布和边际分布都是一元正态分布.

应该注意到在给定 $X_1=x_1$ 的条件下，X_2 的条件分布具有一些特殊性质. 如果 $\rho\neq 0$，则 $\mathrm{E}(X_2|x_1)$ 是给定值 x_1 的线性函数. 如果 $\rho>0$，该线性函数的斜率是正的；如果 $\rho<0$，该线性函数的斜率是负的. 然而，在给定 $X_1=x_1$ 的条件下，X_2 的条件分布的方差是 $(1-\rho^2)\sigma_2^2$，它不依赖于 x_1. 而且，X_2 的条件分布的方差比 X_2 的边际分布的方差 σ_2^2 要小.

例 5.10.2 预测一个人的体重 从一个特定总体中随机抽取一个人，以 X_1 表示此人的身高，X_2 表示此人的体重. 假定 X_1 和 X_2 服从二元正态分布，其概率密度函数由(5.10.4)式确定，并假定需要对 X_2 进行预测. 下面，将此人的身高 X_1 已知时对其体重预测得到的最小均方误差和此人的身高未知时对其体重预测得到的最小均方误差进行比较.

如果此人的身高未知时，则体重的最好预测是均值 $\mathrm{E}(X_2)=\mu_2$，该预测的最小均方误差为方差 σ_2^2. 如果此人的身高已知，且为 x_1，则最好的预测是在给定 $X_1=x_1$ 的条件下 X_2 的条件分布的均值 $\mathrm{E}(X_2|x_1)$，预测的均方误差是条件分布的方差 $(1-\rho^2)\sigma_2^2$. ◀

由于例 5.10.2 中条件分布的方差为 $(1-\rho^2)\sigma_2^2$，所以，不考虑此人的身高 x_1 已知，则对于一个无论个子是高或是矮，还是中等的人来说，预测其体重的难度是一样的. 而且，由于方差 $(1-\rho^2)\sigma_2^2$ 随着 $|\rho|$ 的增加而减少，则当从一个总体随机抽取到的人的身高和体重是高度相关时，很容易由这个人的身高预测其体重.

例 5.10.3 确定边际分布 设随机变量 X 服从均值为 μ、方差为 σ^2 的正态分布，且对于每个 x，另一个随机变量 Y 在给定 $X=x$ 的条件下服从均值为 x、方差为 τ^2 的正态分布. 现在来确定 Y 的边际分布.

我们已知 X 的边际分布是正态分布，在给定 $X=x$ 的条件下，Y 的边际分布也是正态分布，且边际分布的均值是 x 的线性函数，方差是常数. 因此，X 和 Y 的联合分布必定是二元正态分布（见习题 14）. 因此，Y 的边际分布也是正态分布，由此一定可以确定 Y 的均值和方差.

316

Y 的均值为

$$\mathrm{E}(Y)=\mathrm{E}[\mathrm{E}(Y\mid X)]=\mathrm{E}(X)=\mu.$$

并且，由 4.7 节的习题 11，

$$\mathrm{Var}(Y)=\mathrm{E}[\mathrm{Var}(Y\mid X)]+\mathrm{Var}[\mathrm{E}(Y\mid X)]=\mathrm{E}(\tau^2)+\mathrm{Var}(X)=\tau^2+\sigma^2.$$

因此，Y 的边际分布是均值为 μ 且方差为 $\tau^2+\sigma^2$ 的正态分布. ◀

5.10.3 线性组合

设两个随机变量 X_1 和 X_2 服从二元正态分布，其概率密度函数由(5.10.4)式确定. 现在考虑随机变量 $Y=a_1X_1+a_2X_2+b$，其中，a_1、a_2 和 b 是任意给定的常数. X_1 和 X_2 均可由独立的正态随机变量 Z_1 和 Z_2 的线性组合来表示，正如(5.10.2)式的形式. 由于 Y 是 X_1 和 X_2 的线性组合，则 Y 也可以表示成 Z_1 和 Z_2 的线性组合. 因此，根据推论5.6.1，Y 的分布也是正态分布. 因此得到下面重要的性质：

如果两个随机变量 X_1 和 X_2 服从二元正态分布，则每个线性组合 $Y=a_1X_1+a_2X_2+b$ 服从正态分布.

Y 的均值和方差如下：

$$\begin{aligned}\mathrm{E}(Y)&=a_1\mathrm{E}(X_1)+a_2\mathrm{E}(X_2)+b\\&=a_1\mu_1+a_2\mu_2+b\\\mathrm{Var}(Y)&=a_1^2\mathrm{Var}(X_1)+a_2^2\mathrm{Var}(X_2)+2a_1a_2\mathrm{Cov}(X_1,X_2)\\&=a_1^2\sigma_1^2+a_2^2\sigma_2^2+2a_1a_2\rho\sigma_1\sigma_2.\end{aligned} \tag{5.10.8}$$

例5.10.4 丈夫和妻子的身高 假定从一个由已婚夫妇组成的特定总体中随机抽取一对夫妇，设妻子身高和丈夫身高的联合分布为二元正态分布，且妻子身高的均值为66.8英寸，标准差为2英寸，丈夫身高的均值为70英寸，标准差为2英寸，两人身高的相关系数为0.68. 现在来确定妻子比丈夫高的概率.

317

如果以 X 表示妻子的身高，Y 表示丈夫的身高，则需要确定的概率为 $\Pr(X-Y>0)$. 由于 X 和 Y 的联合分布为二元正态分布，所以 $X-Y$ 的分布是正态分布，其均值为

$$\mathrm{E}(X-Y)=66.8-70=-3.2$$

方差为

$$\begin{aligned}\mathrm{Var}(X-Y)&=\mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)\\&=4+4-2(0.68)(2)(2)=2.56.\end{aligned}$$

因此，$X-Y$ 的标准差是1.6.

随机变量 $Z=(X-Y+3.2)/1.6$ 服从标准正态分布. 由本书后面的附录表可以得到：

$$\Pr(X-Y>0)=\Pr(Z>2)=1-\Phi(2)=0.0227.$$

于是，妻子比丈夫高的概率为0.0227. ◀

5.10.4 小结

如果随机向量 (X,Y) 服从二元正态分布，则每个线性组合 $aX+bY+c$ 服从正态分布. 特别地，X 和 Y 的边际分布是正态分布. 在给定 $Y=y$ 的条件下，X 的条件分布也是正态分布，其条件均值为 y 的线性函数，条件方差为常数.（在给定 $X=x$ 条件下，Y 的条件分布也有类似的结论.）在 D. F. Morrison（1990）所著的书中可以找到关于二元正态分布更彻底的讨论和更高维数的一般结论.

习题

1. 再考虑例5.10.4中丈夫和妻子身高的联合分布，求在给定丈夫的身高是72英寸的条件下，妻子身高的条件分布的0.95分位数.
2. 设从某个总体中随机选取一名学生参加两门不同的考试A和B. 假定考试A的成绩平均分为85，标准差为10分，考试B的成绩平均分为90，标准差为16，两门考试成绩的联合分布为二元正态分布，且相关系数为0.8. 如果该学生在考试A中的成绩为80，则在考试B中的成绩高于90分的概率为多少?
3. 再次考虑习题2中有关两门课程的考试. 如果随机抽取一名学生，则她两门课程的总成绩高于200分的概率为多少?

318

4. 再次考虑习题2中有关两门课程的考试. 如果随机抽取一名学生，则在考试A中的成绩高于考试B的成绩的概率为多少?
5. 再次考虑习题2中有关两门课程的考试. 如果随机抽取一名学生，已知她在考试B中的成绩是100分，则她在考试A中成绩的预测值有最小均方误差时的预测值为多少? 最小均方误差为多少?
6. 设随机变量X_1和X_2服从二元正态分布，其联合概率密度函数有(5.10.4)式的形式. 求使得$\text{Var}(X_1+bX_2)$达到最小的常数b.
7. 设随机变量X_1和X_2服从二元正态分布，已知：$\text{E}(X_1|X_2)=3.7-0.15X_2$, $\text{E}(X_2|X_1)=0.4-0.6X_1$, $\text{Var}(X_2|X_1)=3.64$. 求X_1的均值和方差，X_2的均值和方差，X_1和X_2的相关系数.
8. 以$f(x_1,x_2)$表示二元正态分布的联合概率密度函数，具体形式如(5.10.4)式. 证明：$f(x_1,x_2)$的最大值在$x_1=\mu_1$且$x_2=\mu_2$处达到.
9. 以$f(x_1,x_2)$表示二元正态分布的联合概率密度函数，具体形式如(5.10.4)式. 令k是满足下述不等式的常数：
$$0<k<\frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2}.$$
证明：如果$\rho=0$，$\sigma_1=\sigma_2$，则满足$f(x_1,x_2)=k$的点(x_1,x_2)位于一个圆上，否则，位于一个椭圆上.
10. 设随机变量X_1和X_2服从二元正态分布，另两个随机变量Y_1和Y_2具有如下形式：
$$Y_1=a_{11}X_1+a_{12}X_2+b_1,$$
$$Y_2=a_{21}X_1+a_{22}X_2+b_2,$$
其中
$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}\neq 0.$$
证明：Y_1和Y_2也服从二元正态分布.
11. 设随机变量X_1和X_2服从二元正态分布，满足$\text{Var}(X_1)=\text{Var}(X_2)$. 证明：随机变量$X_1+X_2$和$X_1-X_2$相互独立.
12. 考虑例5.10.1，跳甲的两次测量服从二元正态分布，其参数为：$\mu_1=201$，$\mu_2=118$，

$\sigma_1=15.2, \sigma_2=6.6$，$\rho=0.64$. 假定第二类跳甲的两次测量也服从二元正态分布，其参数为：$\mu_1=187$，$\mu_2=131$，$\sigma_1=15.2$，$\sigma_2=6.6$，$\rho=0.64$. 以(X_1, X_2)表示来自这两类中某一类的跳甲的两次测量，令 a_1 和 a_2 是常数.

a. 对于每一类跳甲，求 $a_1X_1+a_2X_2$ 的均值和标准差.（注意到两类跳甲的方差是相同的. 如何判断是相同的?）

b. 求 a_1 和 a_2，使得 a. 中求得的两个均值的差与两个标准差的差的比最大. 即，在所有可能的线性组合中，$a_1X_1+a_2X_2$ 能最好地区分两个物种.

13. 设两个随机变量 X 和 Y 的联合分布与关于(x,y)的下述函数成比例：

$$\exp\left(-[ax^2+by^2+cxy+ex+gy+h]\right),$$

其中 $a>0$，$b>0$，c、e、g 和 h 都为常数. 假定 $ab>(c/2)^2$. 证明：X 和 Y 服从二元正态分布，并求其均值、方差和相关系数.

14. 设随机变量 X 服从正态分布，对于每个 x，另一个随机变量 Y 在给定 $X=x$ 的条件下，服从均值为 $ax+b$、方差为 τ^2 的正态分布，其中 a、b 和 τ^2 是常数. 证明：X 和 Y 的联合分布是二元正态分布.

15. 设随机变量 $X_1, \cdots, X_n$ 独立同分布，且都服从均值为 μ、方差为 σ^2 的正态分布. 定义 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^n X_i$ 为样本均值. 在本题中，我们来求给定 $\overline{X}_n$ 的条件下每个 X_i 的分布.

a. 证明：X_i 和 $\overline{X}_n$ 服从二元正态分布，其均值都为 μ，方差分别为 σ^2 和 σ^2/n，相关系数为 $1/\sqrt{n}$. 提示：令 $Y=\sum_{j\neq i}X_j$，证明：Y 和 X_i 是独立的正态随机变量，且 $\overline{X}_n$ 和 X_i 是 Y 和 X_i 的线性组合.

b. 证明：在给定 $\overline{X}_n=\overline{x}_n$ 的条件下，X_i 的条件分布是均值为 $\overline{x}_n$ 且方差为 $\sigma^2(1-1/n)$ 的正态分布.

319

第6章 估　　计

6.1 统计推断

回顾诊疗试验的例题. 考虑如下问题：即如何利用其他病人的观察结果来对当前病人的治疗是否成功的概率来下论断. 这就是统计推断被用来处理的一类问题. 一般地，统计推断由对未知量的概率推断组成. 例如，我们可以计算与未知变量和分布中的未知参数有关的均值、方差、分位数、概率及其他一些相关的量. 我们的目标就是从收集到的认为可能含有相关信息的数据中可以了解到未知量的什么性质. 我们还有其他一些问题可以用统计推断来解决. 我们观察一台机器的一些产品后应如何对机器是否正常工作下论断；在公民诉讼中，当我们调查了如何判决不同种族群体之后，应怎样对是否有种族歧视来下论断等. 我们将用来解决处理这些问题的统计推断方法是建立在第一章至第五章概率论的基础上的.

6.1.1 统计推断的例题

在本书的前五章中我们讨论了概率论的理论和方法. 如同我们介绍概率论中的新概念一样，对于第六章至第十章的主要内容也介绍一些用概率的理论和方法进行统计推断的例题. 这里，我们将复习一些例题，目的是为了预习一些与大多数统计推断问题相关的基本概念.

323

例 6.1.1 诊所试验 例 2.1.3 中所介绍的诊所试验关注的主要问题是在各种治疗方法下如何使病人避免旧病复发. 例如，我们可以把在丙咪嗪(抗抑郁剂)组中任何一个病人会避免旧病复发的概率 P 看作为一个未知量，它的分布描述了 P 的不确性. 在表 2.1 丙咪嗪列中的病人给我们提供的信息可以降低 P 的不确定性. 统计推论包含了对数据和/或概率 P 的概率描述，以及数据和概率可以相互验证的内容. 例如，在例 2.3.10 中，假定 P 有 11 个可能的值，计算了在给定研究结果条件下，这 11 个值(事件 B_j)的条件概率. 在例 4.7.5 中，为了预测 P，我们在观察研究结果前后分别计算了 M. S. E.. 在例 5.10.3 中，我们计算了在给定的研究结果条件下 P 的条件均值. ◀

例 6.1.2 检查产品 在例题 2.2.9 和 2.3.9 中，一台机器以未知概率生产出次品，但这个概率的分布已知. 当我们抽查了六个产品后，可以在已知六个抽查产品结果的条件下计算出所求未知概率的条件分布. ◀

例 6.1.3 放射性粒子 在例 5.9.1 中，放射性粒子击中目标过程是一个泊松过程，但其速率 λ 未知. 在 5.9 节的习题 22 中，求在观察泊松过程一段时间后计算 λ 的条件分布. ◀

以上这些例题仅说明在得到观察数据后进行统计推断的一些可能的类型. 在本节的其余内容中，我们将更精确地表述哪些内容能被归为统计推断，同时也将解释统计推断中常用的一些术语.

6.1.2 统计推断的本质

一个统计推断问题，或者更简洁地说，一个统计问题就是对以未知概率分布产生的一批数据进行分析，并必须对该未知分布下些论断. 即，有可能会有两个或更多的概率分布会产生这样一批试验数据，而我们要能够确定哪个分布看上去更可能产生这批数据. 在大多数实际问题中，通常可能有无数个不同的分布生成这批数据. 通过数据分析试图了解该未知分布，对未知分布的一些性质下些论断，并确定每个分布实际上是正确的那个分布的相对可能性. 此外，或者进而，也许还希望预测一些未被观察到的未来数据. 由此无论对产生这批已经观察到的数据的分布了解到何种情况都会对预测未来数据有帮助.

324

6.1.3 参数

在很多统计问题中，我们会假定产生试验数据的概率分布已知，除了分布中一个或更多的参数未知. 比如，如 5.9 节中所定义的某种起博器的寿命长度服从参数为 β 的指数分布，而此参数 β 的精确值未知. 如果能够观察到若干个此类起博器的寿命，那么从这些观察数据和其他可能得到的一些相关信息，就有可能对该未知参数 β 作统计推断. 例如，我们可能希望得到 β 的最好估计值，或者是确定 β 可能落在的区间，或者确定 β 是否小于某个给定的值. 但是一般地不可能给出 β 的精确值.

下面来看另一个例子. 假设某个总体中的个体的高度分布服从均值为 μ、方差为 σ^2 的正态分布，其中 μ 和 σ^2 的准确值未知. 如果我们观察到来自母体的一个随机选择的一些个体的高度，那么从这些观察到的高度和与此高度分布相关的其他一些信息，可以对 μ 和 σ^2 的值作推断.

在统计推断问题中，称确定产生实验数据的分布的一个特征或特征的组合为分布的参数. 例如，在刚才的例子中均值 μ 和方差 σ^2 即是用来确定个体身高的正态分布的参数. 参数 θ 或参数向量 $(\theta_1,\cdots,\theta_k)$ 的所有可能取值的集合 Ω 称为参数空间.

在起博器的例子中，指数分布的参数 β 必须是正的. 因此，除非可以确认某个值就是 β 的可能值，参数空间 Ω 就是所有正数的集合. 在高度的例子中，正态分布中的均值 μ 和方差 σ^2 可以被看作是一对参数. 此处 μ 和 σ^2 都必须是正的. 因此参数空间 Ω 可以看作是所有满足 $\mu>0$，$\sigma^2>0$ 的 (μ,σ^2) 的集合. 更特殊地，如果在此例中的正态分布是表示某个特殊的母体中个体的高度(单位为英寸)的分布，我们可以

确定 $30<\mu<100$ 及 $\sigma^2<50$. 这里其相应的样本空间 Ω 应该是由满足 $30<\mu<100$ 且 $0<\sigma^2<50$ 条件的 (μ,σ^2) 组成的一个较小的集合.

参数空间 Ω 的重要特点是它包含了一个给定问题中参数的所有可能取值，以便可以确认参数向量的实际值是样本空间 Ω 中的一个点.

6.1.4　预测

推断的一种形式是预测还未被观测到的随机变量. 例如，可以把为顾客服务的时间看成是一个速率 θ 未知的指数型的随机变量，但是我们可能对 θ 没有直接的兴趣，而是希望能够预测下一个顾客的服务时间或者是为接下来的 k 个顾客所需要的总服务时间. 在诊疗试验的例子中，我们感兴趣的可能是预测在下组丙咪嗪组中有多少病人会治愈. 事实上，在每一个还没有观察到所有相关数据的统计问题中，进行预测是可能的.

325

6.1.5　统计决策问题

在许多统计问题中，对试验数据进行分析后，我们必须在一些可能的决策集合中选取一个，而这些决策的后果依赖于某个参数的未知值. 例如，我们必须要估计未知参数 θ 的值，其后果又依赖于我们的估计值与 θ 的真实值有多接近. 又如另一个例子，需要确定 θ 的未知值是大于还是小于某个常数，而后果又依赖于我们的推断是否正确.

6.1.6　试验设计

在一些统计问题中，我们要控制采集的试验数据的类型或数量. 例如，把一个确定某种合金平均拉力强度的试验看作是合金合成时压力和温度的函数. 在一定的预算和时间限制内，试验者可以选择在怎样的压力和温度的水平下生产合金，而且可以在不同的水平下确定样本的生产数量.

试验者可以(至少在某种程度)选择实施试验方案的问题称为试验设计问题. 当然，试验设计和对试验数据的统计分析是息息相关的. 如果不考虑对将要得到的数据进行统计分析，那么就不可能设计出一个有效的试验；同样，如果不考虑取得数据的试验的特定类型，也不可能对试验数据进行有意义的统计分析.

6.1.7　参考文献

在本书的其余部分，我们将考虑统计推断、统计判决和试验设计中许多不同的问题. 有些关于统计理论和方法的书与本书1.1节中所提到的内容程度一样. 有些程度更高的关于统计的书有 Bickel 和 Doksum(2000)，Casella 和 Berger(2002)，Cramer(1946)，DeGroot(1970)，Ferguson(1967)，Lehmann(1997)，Lehmann 和 Casella(1998)，Rao(1973)，Rohatgi(1976)和 Schervish(1995).

326

6.2 先验分布与后验分布

在得到任何观察数据之前参数的分布称为参数的先验分布. 在给定观察数据条件下参数的分布称为后验分布. 如果我们把数据的观察值代入给定参数下数据的条件概率函数或者概率分布函数中, 那么结果仅是参数的函数, 即称为似然函数.

6.2.1 先验分布

在本书 2.3 节, 定义事件的先验概率是在收集当前所关注的数据之前给定那些事件的概率. 我们已经看到了一些例子如何把这一思想扩展到统计模型中参数的先验分布中去. 例如, 在例 3.4.2 中, 参数 P 有 11 个不同的取值, 我们所用的先验分布为取每个值的相应概率为 1/11. 在 5.9 节的习题 22 中, 参数 λ 的先验分布是参数为 α 和 β 的 gamma 分布.

1. 确定先验分布

考虑一个从概率密度函数或概率函数为 $f(x\mid\theta)$ 的分布中所得观察值的统计推断问题, 其中参数 θ 的值未知. 假定参数 θ 的未知值落在某个特定的参数空间 Ω 中. 那么在 6.1 节中介绍的一些统计推断问题可以大致地描述为以从概率密度函数或概率函数 $f(x\mid\theta)$ 中所得的观察值为基础, 确定参数 θ 的真值究竟落在参数空间 Ω 的何处.

在很多问题中, 试验者或统计学家在从 $f(x\mid\theta)$ 得到任何观察值之前, 通过构造 θ 在集合 Ω 中的一个概率分布来归纳有关 θ 在集合 Ω 中位置的先验的信息和知识. 即, 在收集或观察到任何试验数据之前, 试验者利用已有的经验和知识可以使他判断 θ 更有可能落在 Ω 的某个范围内, 而不是 Ω 的其他区域. 假设不同区域的相对可能性可以用 Ω 上的概率分布来表示, 称这个分布为 θ 的**先验分布**, 因为该分布描述了从 $f(x\mid\theta)$ 中得到任何观察值之前 θ 的真值落在 Ω 中不同区域的相对可能性.

注: 随机样本和有条件的独立. 我们用来表示数据的概率函数或者概率密度函数的记号 $f(x\mid\theta)$ 看上去像是在给定 θ 值的条件下的条件概率函数或概率密度函数. 如果把 θ 看作是一个随机变量, 那么 $f(x\mid\theta)$ 事实上是每个观察值 X_i 在给定 θ 值下的条件概率函数或概率密度函数. 在该种情况中, 以下两种表示是等价的:

- $X_1,\cdots,X_n$ 组成一个随机样本, 其概率函数或概率密度函数为 $f(x\mid\theta)$.
- $X_1,\cdots,X_n$ 是在给定 θ 条件下条件 i. i. d. , 其条件概率函数或者概率密度函数为 $f(x\mid\theta)$.

327

虽然为了简便, 在一般情况下会用以上的第一种表示, 但是当把 θ 看作是一个随机变量时, 记住这两种表示等价性是有用的.

2. 先验分布的争论本质

先验分布的概念在统计中非常有争议. 该争议与我们在 1.2 节中讨论的关于

概率的意义的争议有密切关系. 一些统计学者相信在每一个统计问题中对于参数 θ 都可以选定一个先验分布. 他们认为该分布是一个主观的概率分布，而这是由于它所表示的是试验者个人的信息和对于真值 θ 落在何处的主观臆断. 他们还认为，无论如何先验分布和在统计领域使用的其他概率分布没有区别，而且概率论中所有的准则都适用于先验分布. 称这些统计学家是支持统计中的贝叶斯准则的.

另外一些统计学家认为在许多问题中谈及 θ 的概率分布是不合适的. 因为 θ 根本就不是一个随机变量，而是某个固定的值，而仅仅对于试验者而言是未知的. 这些统计学家认为只有当掌握 θ 以前的每个可能取值的相对频数才可以给 θ 指定一个先验分布. 在此情况下，两种不同观点的科学家在对所使用的正确的先验分布上持有相同观点. 例如，在一大批生产的产品中次品的比率 θ 是未知的，假定同一生产商在过去也生产过多批这样的产品，而且保留有过去产品中次品的比率的详细记录. 可以用过去多批产品中的相对频率来建立 θ 的先验分布.

两组统计学家一致同意如果可以选定一个有意义的先验分布的话，那么在本节中所描述的理论和方法都是适用和有用的. 在本节和6.3节中，我们将假定可以给 θ 指定一个先验分布，表示未知参数 θ 落在参数空间不同子集中的概率. 在6.4节的开始部分，将考虑一些不基于指定先验分布条件下的估计技巧.

3. 不适当的先验分布

在2.3节例2.3.11中，两组截然不同的先验概率被用在同一系列事件上. 然而当我们观察到数据以后，发现后验概率很相似. 这是非常常见的情形，在观察到数据之后，不同的先验分布差别并不大. 特别是当数据很多或是被比较的先验分布十分分散时尤为如此. 该事实有两个含义. 第一，不同的试验者有可能在先验分布上会有分歧意见，但事实是当我们有很多数据时，此分歧就变得无关紧要了. 第二，当指定哪个先验分布都没有很大关系时，试验者就不会倾向于花费很多时间来指定先验分布. 但不幸的是，如果不指定先验分布，就没有办法算出给定数据条件下参数的条件分布. 作为一种尝试，某些计算似乎显示数据比可用的先验分布含有更多的信息. 通常地，这些计算涉及一个函数 $\xi(\theta)$，把它当作参数 θ 的先验概率密度函数，但是这里 $\int \xi(\theta)\,d\theta=\infty$，而这明显的与概率密度函数的定义相违背. 称这样的先验为不适当的.

328

4. 离散的和连续的先验分布

在一些问题中，参数 θ 只可能取有限个不同值或至多是取不同值的一个无限序列(例题2.3.10和2.3.11就是这种类型). 因此 θ 的先验分布是一个离散型分布. 称这个分布的概率函数 $\xi(\theta)$ 为 θ 的先验概率函数. 在另一些问题中，参数 θ 可以取实数轴上任意的值，或取值于实数轴上的一个区间，那么就对 θ 设定一个连续型的先验分布. 称该分布的概率密度函数 $\xi(\theta)$ 为 θ 的先验概率密度函数.

例 6.2.1 一枚普通的或两面人头的硬币 令 θ 表示掷一枚硬币时得到人头的概率，假定硬币可能是普通的或两面都是人头，因此 θ 的可能取值只能是 $\theta=1/2$ 和 $\theta=1$. 如果硬币是普通的硬币其先验概率为 p，那么 θ 的先验概率函数为 $\xi(1/2)=p$ 和 $\xi(1)=1-p$. ◀

例 6.2.2 次品的比例 假设一批产品中次品的比例 θ 是未知的，且给 θ 指定的先验分布是区间[0,1]上的均匀分布. 那么 θ 的先验概率密度函数是

$$\xi(\theta)=\begin{cases}1 & 对于 0<\theta<1,\\ 0 & 其他.\end{cases} \tag{6.2.1}$$ ◀

例 6.2.3 指数分布的参数 假设某种类型荧光灯的寿命(以小时计)是待观察的，而且任意一盏灯的寿命时间的分布是如 5.9 节中定义的一个参数为 β 的指数分布. 假定 β 的确切值是未知的，但根据以前的经验 β 的先验分布可以取为均值为 0.0002、标准差为 0.0001 的伽玛分布. 我们要确定 β 的先验概率密度函数.

假设 β 的先验分布是参数为 α_0 和 β_0 的伽玛分布. 在 5.9 节中已证明该分布的均值为 α_0/β_0 以及其方差为 α_0/β_0^2. 因此 $\alpha_0/\beta_0=0.0002$ 和 $\alpha_0^{1/2}/\beta_0=0.0001$. 因而可以求得 $\alpha_0=4$ 和 $\beta_0=20\,000$. 由方程(5.9.7)可以得到当 $\beta>0$ 时，β 的先验概率函数为

$$\xi(\beta)=\frac{(20\,000)^4}{3!}\beta^3\mathrm{e}^{-20\,000\beta}. \tag{6.2.2}$$

且当 $\beta\leqslant 0$ 时，$\xi(\beta)=0$. ◀

6.2.2 后验分布

称给定观察值条件下参数 θ 的条件分布为 θ 的后验分布. 对于(3.6.13)中的随机变量和(3.7.14)中的随机向量的贝叶斯定理告诉我们当得到观察值后如何来计算 θ 后验分布的概率密度函数或概率函数. 我们将用先验分布和参数中特定的记号来复习贝叶斯定理的推导. 329

假设有一个来自于概率密度函数或概率函数为 $f(x|\theta)$ 的容量为 n 的随机样本 $X_1,\cdots,X_n$. 同样假定参数 θ 的值未知，而且 θ 的先验概率密度函数或先验概率函数为 $\xi(\theta)$. 为简便起见，假定参数空间 Ω 是实数轴上的一个区间或者是整个实数轴，$\xi(\theta)$ 是 Ω 上的一个先验概率密度函数，而不是先验概率函数，$f(x|\theta)$ 是概率密度函数而不是概率函数. 但是以下的讨论对于 $\xi(\theta)$ 或者 $f(x|\theta)$ 是一个概率函数时也适用.

由于随机变量 $X_1,\cdots,X_n$ 是来自于概率密度函数为 $f(x|\theta)$ 的分布的随机样本，由 3.7 节可知其联合概率密度函数 $f_n(x_1,\cdots,x_n|\theta)$ 由以下等式给出

$$f_n(x_1,\cdots,x_n|\theta)=f(x_1|\theta)\cdots f(x_n|\theta). \tag{6.2.3}$$

如果用向量记号 $\boldsymbol{x}=(x_1,\cdots,x_n)$，那么(6.2.3)式中的联合概率密度函数可以简记为 $f_n(\boldsymbol{x}|\theta)$.

既然认为参数 θ 本身具有概率密度函数为 $\xi(\theta)$ 的分布，那么其联合概率密度函数 $f_n(\boldsymbol{x}|\theta)$ 可以被看作是在给定 θ 值条件下 $X_1,\cdots,X_n$ 的条件联合概率密度函数. (6.2.3)式仅表明 $X_1,\cdots,X_n$ 是条件独立的且在给定 θ 条件下是同分布的，其概率密度函数都是 $f(x|\theta)$. 如果条件联合密度函数乘以概率密度函数 $\xi(\theta)$，可得 $X_1,\cdots,X_n$ 和 θ 的 $(n+1)$ – 维联合概率密度函数为 $f_n(\boldsymbol{x}|\theta)\xi(\theta)$. $X_1,\cdots,X_n$ 的边际联合概率密度函数可以通过对这个联合概率密度函数对于 θ 的所有取值的积分来得到. 所以，可以将 $X_1,\cdots,X_n$ 的 n – 维边际联合概率密度函数表示成如下形式

$$g_n(\boldsymbol{x}) = \int_\Omega f_n(\boldsymbol{x}|\theta)\xi(\theta)\,\mathrm{d}\theta. \tag{6.2.4}$$

(6.2.4)式是随机向量(3.7.13)的全概率公式的一个例子.

进而，在给定 $X_1=x_1,\cdots,X_n=x_n$ 条件下 θ 的条件概率密度函数，用 $\xi(\theta|x)$ 表示，必须等于 $X_1,\cdots,X_n$ 和 θ 的联合概率密度函数除以 $X_1,\cdots,X_n$ 的联合边际概率密度函数. 因而有

$$\xi(\theta|\boldsymbol{x}) = \frac{f_n(\boldsymbol{x}|\theta)\xi(\theta)}{g_n(\boldsymbol{x})}, \text{对所有 } \theta\in\Omega. \tag{6.2.5}$$

以上是对于参数和随机样本重述的贝叶斯定理.

(6.2.5)式中由条件概率密度函数所表示的 Ω 上的概率分布叫作 θ 的*后验分布*，因为这是在观察到 $X_1,\cdots,X_n$ 的值后 θ 的分布. 类似地，称(6.2.5)式中 θ 的条件概率分布为 θ 的*后验概率密度函数*. 事实上，我们可以说先验概率密度函数 $\xi(\theta)$ 表示的是在得到观察值之前 θ 的真值落在 Ω 中各种区域的相对可能性；而后验概率密度函数 $\xi(\theta|x)$ 则表示在观察到 $X_1=x_1,\cdots,X_n=x_n$ 的值后的相对可能性.

330

6.2.3 似然函数

(6.2.5)式右边的分母是分子对于 θ 所有可能取值的一个积分. 虽然该积分的值依赖于观察值 $x_1,\cdots,x_n$，但是它的确不依赖于 θ，所以将(6.2.5)式的右边看作 θ 的一个概率密度函数时可以认为它是一个常数. 因而(6.2.5)式可以用下列关系来替代

$$\xi(\theta|\boldsymbol{x}) \propto f_n(\boldsymbol{x}|\theta)\xi(\theta). \tag{6.2.6}$$

这里用的比例符号 $\propto$ 表示左式和右式除了相差一个依赖于观察值 $x_1,\cdots,x_n$，但不依赖于 θ 的一个常因子外是相等的. 这个使(6.2.6)式中两边相等的常数因子在任何时候都可以用如下事实来决定 $\int_\Omega \xi(\theta|\boldsymbol{x})\,\mathrm{d}\theta = 1$，因为这里的 $\xi(\theta|\boldsymbol{x})$ 是 θ 的概率密度函数.

当把一个随机样本中的观察值的联合概率密度函数或者联合概率分布 $f_n(\boldsymbol{x}|\theta)$ 看作是在给定观察值 $x_1,\cdots,x_n$ 的条件下 θ 的函数，那么就把它称为*似然函数*. 在这个术语中，(6.2.6)式表明了 θ 的后验分布与似然函数和 θ 的先验分布的乘积成正比.

通过(6.2.6)中的比例关系式，我们一般不用对(6.2.4)式进行精确的积分而可以确定 θ 的后验概率密度函数. 如果我们识别出关系式(6.2.6)式的右边可能除了一个常数因子外恰好等于第五章中所介绍的某个标准的或是本书中其他地方的某个概

率密度函数，那么可以非常容易地确定这个合适的因子，从而把(6.2.6)式的右边变成θ的一个正确的概率密度函数，我们通过例6.2.2和6.2.3来进一步阐述这一观点.

例6.2.4 次品的比例 如例6.2.2中的假设，在一批产品中次品的比例θ未知，而且θ的先验分布是区间[0,1]上的一个均匀分布. 假设从一批产品中抽取容量为n的一个随机样本，并且对于$i=1,\cdots,n$，令$X_i=1$表示第i个产品是次品，否则$X_i=0$. 则$X_1,\cdots,X_n$形成一个参数为θ的n重伯努利试验. 我们要确定θ的后验概率密度函数.

从(5.2.2)式可以得到每个观察值X_i的概率分布为

$$f(x|\theta)=\begin{cases}\theta^x(1-\theta)^{1-x} & \text{对于 } x=0.1,\\ 0 & \text{其他.}\end{cases}$$

因而，如果令$y=\sum_{i=1}^n x_i$，那么$X_1,\cdots,X_n$的联合概率函数可以表达为如下形式，对于$x_i=0$或$1(i=1,\cdots,n)$:

$$f_n(\boldsymbol{x}|\theta)=\theta^y(1-\theta)^{n-y}. \tag{6.2.7}$$

由于先验概率密度函数可以由方程(6.2.1)式给定，因此当$0<\theta<1$时，有

$$f_n(\boldsymbol{x}|\theta)\xi(\theta)=\theta^y(1-\theta)^{n-y}. \tag{6.2.8}$$

把上述表达式与(5.10.1)式作比较可以看到，除了一个常数因子外，它是一个参数为$\alpha=y+1$和$\beta=n-y+1$的β分布的概率密度函数. 由于后验概率密度函数$\xi(\theta|x)$和(6.2.8)式的右边成比例关系，因而$\xi(\theta|\boldsymbol{x})$一定是参数为$\alpha=y+1$和$\beta=n-y+1$的$\beta$分布的概率密度函数. 因此，当$0<\theta<1$时，

331

$$\xi(\theta|\boldsymbol{x})=\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y}. \tag{6.2.9}$$ ◀

例6.2.5 一个指数分布的参数 如例6.2.3一样，假定某种型号的荧光灯的寿命分布为一个参数为β的指数分布，而且β的先验分布是如(6.2.2)式中给定的概率密度函数为$\xi(\beta)$的一个特殊的伽玛分布. 假定观察到该种类型样本容量为n的随机样本$X_1,\cdots,X_n$的观察值，要确定在给定$X_1=x_1,\cdots,X_n=x_n$条件下β的后验概率密度函数.

由(5.9.10)式可知，每个观察值X_i的概率密度函数为

$$f(x|\beta)=\begin{cases}\beta e^{-\beta x} & \text{当 } x>0,\\ 0 & \text{其他.}\end{cases}$$

因此，如果我们令$y=\sum_{i=1}^n x_i$，则$X_1,\cdots,X_n$的联合概率密度函数就可以表示成如下形式：当$x_i>0$时，

$$f_n(\boldsymbol{x}|\beta)=\beta^n e^{-\beta y}. \tag{6.2.10}$$

由于先验概率密度函数是由(6.2.2)式给定的，因而对于$\beta>0$，有

$$f_n(\boldsymbol{x}|\beta)\xi(\beta)\propto\beta^{n+3}e^{-(y+20\,000)\beta}. \tag{6.2.11}$$

在关系式(6.2.11)的右边省略了一个与β无关的常数因子.

把上述表达式与方程(5.9.7)作比较可以看到，仅除了一个常数因子，它是参数为$n+4$和$y+20\,000$的伽玛分布. 因为后验概率密度函数$\xi(\beta|\boldsymbol{x})$与$f_n(\boldsymbol{x}|\beta)\xi(\beta)$成比例，所以$\xi(\beta|\boldsymbol{x})$一定是一个参数为$n+4$和$y+20\,000$的伽玛分布的概率密度函数. 因此，对于$\beta>0$，有

$$\xi(\beta|\boldsymbol{x})=\frac{(y+20\,000)^{n+4}}{(n+3)!}\beta^{n+3}e^{-(y+20\,000)\beta}. \tag{6.2.12}$$ ◀

注： *对后验概率密度函数的规范化常数.* 从(6.2.11)到(6.2.12)式的过程是用来确定后验概率密度函数的一个常用技术的示例. 可以把任何带来不方便的常数因子从概率密度函数和如同(6.2.6)式中把它们连乘起来的似然函数中去除掉. 然后考察一下乘积的结果，称其为$\xi(\theta)$，判别一下它是否是我们在别处见过的概率密度函数中的一部分. 事实上，如果能找到一个已命名的分布的概率密度函数等于$cg(\theta)$，那么我们的后验概率密度函数也是$cg(\theta)$，而且后验分布的名称也同它一样，就如例题6.2.4和6.2.5所示的一样.

332

6.2.4　小结

参数的先验分布描述了在观察到任何数据前参数的不确定性. 当把似然函数看作是参数的一个函数时，把观察数据代入，那么似然函数就是给定参数条件下数据的条件概率密度函数或概率函数. 该似然函数告诉我们数据可以改变多少不确定性. 似然函数大的值对应的是后验概率密度函数或者概率函数要比先验分布数值大的参数值. 似然函数小的值会出现在后验要比先验数值小的参数值情形中. 参数的后验分布就是在给定数据条件下参数的条件分布. 这可以由首次在3.6节中看到的随机变量的贝叶斯定理来得到. 进而可以用在3.7节中看到的全概率公式的条件形式，预测在给定θ条件下与已有观察值条件独立的未来观察值.

习题

1. 重新考虑例6.2.6中的情形. 假定试验者认为β的先验分布是一个参数为1和5000的伽玛分布. 试验者计算出$\Pr(X_6>3000|\boldsymbol{x})$的值是什么?
2. 假定已知在一大批产品中次品的比例θ的值不是0.1，就是0.2，而且θ的先验概率分布为：

$$\xi(0.1)=0.7\text{ 和 }\xi(0.2)=0.3.$$

并假定当从这批产品中随机抽取8个产品，发现恰好有2个次品. 试决定θ的后验概率

函数.

3. 假定一卷磁带的失效处的数目服从一个泊松分布，其均值 λ 的值要么为1.0，要么为1.5，而且 λ 的先验概率分布如下：

$$\xi(1.0)=0.4 \text{ 和 } \xi(1.5)=0.6.$$

如果随机抽取一卷磁带，发现有三处失效，那么 λ 的后验概率分布是什么？ [334]

4. 假定某个参数 θ 的先验分布是一个均值为10以及方差为5的伽玛分布. 试决定 θ 的先验概率密度函数.

5. 假定某个参数 θ 的先验分布是一个均值为1/3、方差为1/45的 β 分布，试确定 θ 的先验概率密度函数.

6. 假设一大批产品中次品的比例 θ 是未知的，且 θ 的先验分布是区间[0,1]上的均匀分布. 从这批产品中随机地抽取8个产品，发现恰好有3个是次品. 试确定 θ 的后验分布.

7. 与习题6中所描述的情形相似，但假定 θ 的先验概率密度函数为：

$$\xi(\theta)=\begin{cases}2(1-\theta) & \text{当 } 0<\theta<1, \\ 0 & \text{其他.}\end{cases}$$

与习题6的条件相同，假定在一个容量为8的随机样本中发现恰好有3个是次品. 试确定 θ 的后验分布.

8. 假设 $X_1,\cdots,X_n$ 是来自于概率密度函数为 $f(x|\theta)$ 的分布的一个随机样本，其中 θ 的值未知，且 θ 的先验概率密度函数为 $\xi(\theta)$. 证明：由方程(6.2.5)式直接计算得到的后验分布与用(6.2.13)、(6.2.14)和(6.2.15)式逐步得到的结果是相同的.

9. 再次考虑习题6中的问题，假定 θ 的先验分布和习题6中一样，只不过现在假定不从产品中随机抽取一个样本容量为8的样本，而是进行如下试验：从这批产品中一个接着一个随机抽取产品，直到恰好发现3个次品终止. 如果发现在该试验中，我们必须要抽取容量为8的产品，那么试验结束时 θ 的后验分布是什么？

10. 假定从区间 $\left[\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right]$ 上的一个均匀分布中取一个观察值，θ 的值未知，且 θ 的先验分布是区间[10,20]上的均匀分布. 如果 X 的观察值是12，则 θ 的后验分布是什么？

11. 考虑习题10中的条件，而且假定 θ 的先验分布相同. 现在如果从区间 $\left[\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right]$ 上的一个均匀分布中取6个观察值，它们的值分别为11.0，11.5，11.7，11.1，11.4和10.9. 试确定 θ 的后验分布. [335]

6.3 贝叶斯估计

一个参数的估计是数据的某一个函数，希望它接近于参数的真值. 贝叶斯估计是度量参数之间接近程度的某些测度，例如均方误差或者绝对误差下使后验均值最小化的一个估计.

6.3.1 估计问题的本质

在例6.3.5中，我们发现θ的后验分布，测量到的和公布的卡路里数间的平均差别. 一组消费者希望无需明确指定θ的整个分布，而公布单个值来作为θ的估计即可. 那么一般如何来选择这样一个单值估计就是本节的主题.

假设随机变量$X_1,\cdots,X_n$是来自于概率函数或概率密度函数为$f(x|\theta)$的分布的一个随机样本，其中参数θ的值是未知的. 同样假设θ的值必须落入实数轴上某个给定区间Ω内. 该区间Ω可以是有界的，也可以是无界的，特别地，甚至可以是整个实数轴. 最后，假设θ的值必须通过观察到的样本值来估计.

346

基于随机变量$X_1,\cdots,X_n$的参数估计是一个实值函数$\delta(X_1,\cdots,X_n)$，它通过$X_1,\cdots,X_n$的每组可能的值来确定θ的估计值. 即，若$X_1,\cdots,X_n$的观察值是$x_1,\cdots,x_n$，则θ的估计值就是$\delta(x_1,\cdots,x_n)$. 由于θ的值必须属于区间Ω，那么要求估计$\delta(X_1,\cdots,X_n)$的每个可能取值都必须属于Ω是合理的.

要区分估计量和估计这两个概念十分简便. 因为估计量$\delta(X_1,\cdots,X_n)$是随机变量$X_1,\cdots,X_n$的一个函数，估计量本身是一个随机变量，而且它的概率分布可以由$X_1,\cdots,X_n$的联合分布来得到. 另一方面，估计值是通过把特定的观察值$x_1,\cdots,x_n$代入所确定的特定值$\delta(x_1,\cdots,x_n)$. 因此为了方便使用矢量符号$\boldsymbol{X}=(X_1,\cdots,X_n)$和$\boldsymbol{x}=(x_1,\cdots,x_n)$表示. 在该符号中，一个估计是随机向量$\boldsymbol{X}$的一个函数$\delta(\boldsymbol{X})$，而一个估计值就是特定的$\delta(\boldsymbol{x})$的值. 通常将记号$\delta(\boldsymbol{X})$简记为$\delta$要更为简便一些.

6.3.2 损失函数

对于好的估计量δ的最主要的要求就是它要产生一个估计，而产生θ的估计值要接近于θ的真实值. 即一个好的估计量要求误差$\delta(\boldsymbol{X})-\theta$接近于0的概率相当大. 我们假定对于$\theta$的每个可能取值和每个可能属于区间$\Omega$的估计值$a$，当参数的真值为$\theta$，估计值是$a$时，用一个数$L(\theta,a)$来度量统计者的损失或成本. 特别地，当$a$和$\theta$相差得越大，$L(\theta,a)$的数值就越大.

与前面相同，令$\xi(\theta)$表示区间Ω上的先验概率密度函数，考虑这样一个问题，统计者必须在没有观察到随机样本的值的条件下来估计θ的值. 如果统计者选用一个特定的估计值a，那么她的期望损失将为：

$$\mathrm{E}[L(\theta,a)]=\int_\Omega L(\theta,a)\xi(\theta)\,\mathrm{d}\theta. \tag{6.3.1}$$

我们假定统计者希望选取的估计值a使得上述的(6.3.1)式中的期望损失达到最小. 在一个估计问题中，称一个使期望$\mathrm{E}[L(\theta,a)]$最小化的函数L为**损失函数**.

6.3.3 贝叶斯估计量的定义

假设统计者在估计θ前就观察到随机向量$\boldsymbol{X}$的值$\boldsymbol{x}$，并且令$\xi(\theta|\boldsymbol{x})$表示$\theta$在区间

Ω 上的后验概率密度函数. 对于统计者所可能用的每一个估计值 a，她在该情形下的期望损失为

$$E[L(\theta,a)|x] = \int_{\Omega} L(\theta,a)\xi(\theta|x)\,d\theta. \tag{6.3.2}$$

因此，统计者应该要选取一个估计值 a，使等式(6.3.2)中的期望达到最小值. 347

对于随机向量 $\boldsymbol{X}$ 的每一个可能值 $\boldsymbol{x}$，令 $\delta^*(\boldsymbol{x})$ 表示使等式(6.3.2)中的期望损失达到最小的 a 的估计值，那么用该方法来确定值的函数 $\delta^*(\boldsymbol{X})$ 将是 θ 的一个估计值. 该估计量称为是 θ 的贝叶斯估计量. 即对于 $\boldsymbol{X}$ 的每个可能取值 $\boldsymbol{x}$，它的贝叶斯估计的值 $\delta^*(\boldsymbol{x})$ 由下式确定：

$$E[L(\theta,\delta^*(\boldsymbol{x}))|\boldsymbol{x}] = \min_{a\in\Omega} E[L(\theta,a)|\boldsymbol{x}]. \tag{6.3.3}$$

综上，我们已经考虑一个来自于包含参数 θ 的分布的随机样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的估计问题，其中 θ 属于某个指定区间 Ω，但其值未知. 对于每一个给定的损失函数 $L(\theta,a)$ 和每个先验概率分布函数 $\xi(\theta)$，θ 的贝叶斯估计就是对于每个 $\boldsymbol{X}$ 的可能取值 $\boldsymbol{x}$ 都满足等式(6.3.3)的估计量. 应该强调的是，贝叶斯估计的形式将取决于问题中使用的损失函数和指定的 θ 的先验分布.

6.3.4 不同的损失函数

1. 平方误差损失函数

显然，在估计问题中使用最普遍的损失函数是平方误差损失函数. 该函数的定义如下：

$$L(\theta,a) = (\theta - a)^2. \tag{6.3.4}$$

在使用平方误差损失函数时，对于 $\boldsymbol{x}$ 的每一个观察值的贝叶斯估计就是使得期望 $E[(\theta-a)^2|\boldsymbol{x}]$ 最小的 a 的值.

定理4.7.2指出，由 θ 的后验分布计算出 $(\theta-a)^2$ 的期望时，当 a 的值等于后验分布的均值 $E(\theta|\boldsymbol{x})$ 时，该期望将是最小的. 此结论表明在使用(6.3.4)式中的平方误差损失函数时，贝叶斯估计为 $\delta^*(\boldsymbol{X}) = E(\theta|\boldsymbol{X})$.

例6.3.1 估计伯努利分布的参数 设有一个来自于参数 θ 未知的伯努利分布的随机样本 $X_1,\cdots,X_n$，要估计 θ，θ 的先验分布是给定参数为 α 和 β（$\alpha>0$，$\beta>0$）的 β 分布. 我们使用等式(6.3.4)中所定义的平方误差损失函数，此处 $0<\theta<1$ 且 $0<\alpha<1$. 要确定 θ 的贝叶斯估计.

对于观察值 $x_1,\cdots,x_n$，令 $y=\sum_{i=1}^{n} x_i$. 那么由定理6.3.1可知 θ 的先验分布将是参数为 $\alpha+y$ 和 $\beta+n-y$ 的 β 分布. 由参数为 α_1 和 β_1 的 β 分布的均值为 $\alpha_1/(\alpha_1+\beta_1)$，而这个 θ 的后验分布的均值为 $(\alpha+y)/(\alpha+\beta+n)$. 而对于每一个观察到的向量 $\boldsymbol{x}$，其贝叶斯估计 $\delta(\boldsymbol{x})$ 就等于该值. 因此，贝叶斯估计 $\delta^*(\boldsymbol{X})$ 具有如下形式： 348

$$\delta^*(\boldsymbol{X}) = \frac{\alpha + \sum_{i=1}^{n} X_i}{\alpha + \beta + n}. \tag{6.3.5}$$ ◀

例 6.3.2 估计正态分布的均值 设有一个来自于均值 θ 未知而方差 σ^2 已知的正态分布的随机样本 $X_1,\cdots,X_n$. 假定 θ 的先验分布是给定均值 μ 和方差 v^2 的值的正态分布. 最后假设用等式(6.3.4)中定义的平方误差损失函数（其中 $-\infty<\theta<\infty$, $-\infty<a<\infty$）来确定 θ 的贝叶斯估计.

由相关定理可知对于所有的观察值 $x_1,\cdots,x_n$, θ 的后验分布将是一个正态分布，其均值可确定. 因此，贝叶斯估计 $\delta^*(\boldsymbol{X})$ 具有如下形式：

$$\delta^*(\boldsymbol{X})=\frac{\sigma^2\mu+nv^2\overline{X}_n}{\sigma^2+nv^2}. \tag{6.3.6}$$ ◀

2. 绝对误差损失函数

在估计问题中另一常用的损失函数就是绝对误差损失函数. 其定义如下：

$$L(\theta,a)=|\theta-a|. \tag{6.3.7}$$

对于 $\boldsymbol{x}$ 的每一个观察到的值，其贝叶斯估计 $\delta^*(\boldsymbol{x})$ 是使得期望 $E(|\theta-a|\,|\,\boldsymbol{x})$ 达到最小的 a 值.

在 4.5 节中已证明了，对于 θ 的每一个给定的概率分布，当 a 等于 θ 的分布的中位数时，$|\theta-a|$ 的期望将达到最小. 因此，当相对于 θ 的后验分布计算 $|\theta-a|$ 的期望时，a 的取值是 θ 的后验分布的中位数时，该期望才会是最小的. 据此，使用绝对误差损失函数时，得到的贝叶斯估计量 $\delta^*(\boldsymbol{X})$ 的值总是等于 θ 的后验分布的中位数. 我们现在再次考虑例 6.3.1 和例 6.3.2，但这次我们将使用绝对误差损失函数而不是平方误差损失函数.

例 6.3.3 估计伯努利分布的参数 再次考虑例 6.3.1 的情况，但现在要使用式(6.3.7)所指定的绝对误差损失函数. 对于所有的观察值 $x_1,\cdots,x_n$, 其贝叶斯估计 $\delta^*(\boldsymbol{x})$ 将是 θ 的后验分布为参数为 $\alpha+y$, $\beta+n-y$ 的 β 分布的中位数. 这里没有关于该中位数的简单表达式. 它必须通过对于每个给定观察值的集合的数值逼近来确定. 大多数的统计计算机软件都可以计算出任意的 β 分布的中位数. ◀

349

例 6.3.4 正态分布的均值估计 再一次考虑例 6.3.1 中的条件，但现在假定如等式(6.3.7)所指定的绝对误差损失函数，对于所有观察值 $x_1,\cdots,x_n$, $\delta^*(\boldsymbol{x})$ 的贝叶斯估计将等于 θ 的后验正态分布的中位数. 然而，既然每个正态分布的均值和中位数都是相等的，那么 $\delta^*(\boldsymbol{x})$ 就等于后验分布的均值. 因此关于绝对误差损失函数所得到的贝叶斯估计和在平均误差损失函数条件下得到的贝叶斯估计是一样的，它由等式(6.3.6)给出. ◀

3. 其他损失函数

虽然平方误差损失函数和使用范围要小一些的绝对误差损失函数都是估计问题中最为普遍使用的，但是这些损失函数可能在某个特定问题中都不适用. 在此类问题中，可能要用 $L(\theta,a)=|\theta-a|^k$ 这种形式的损失函数才适合，其中 k 是除了 1 或 2 的某个正数. 在其他问题中，当误差 $|\theta-a|$ 有一个指定的量时，损失的结果就有可

能依赖于 θ 的真值. 在此种问题中，就适合用 $L(\theta,a)=\lambda(\theta)(\theta-a)^2$ 或者 $L(\theta,a)=\lambda(\theta)|\theta-a|$ 这类形式的损失函数，其中 $\lambda(\theta)$ 是 θ 的给定的一个正函数. 在另外一些问题中，对于 θ 高估某一定量的值比低估 θ 同样量的值的代价要更大一些. 能够反映该性质的一个特别的损失函数如下：

$$L(\theta,a)=\begin{cases}3(\theta-a)^2 & 若\ \theta\leqslant a,\\ (\theta-a)^2 & 若\ \theta>a.\end{cases}$$

其他各种形式的损失函数可能与特定的估计问题相关. 不过，在本书中我们将大部分集中于平方误差和绝对误差损失函数.

6.3.5 对大样本的贝叶斯估计

1. 不同先验分布的作用

假设在一批大宗运输货物中的次品比例 θ 是未知的，而且 θ 的先验分布是区间 $[0,1]$ 上的均匀分布. 假设必须要估计 θ 的值，而且使用平均误差损失函数. 最后假定从运输的货物中抽取 100 个随机样本，恰好有 10 个是次品. 由于均匀分布是一个参数为 $\alpha=1$ 和 $\beta=1$ 的 β 分布，而且在给定样本中 $n=100, y=100$，根据式(6.3.5)可知，贝叶斯估计是 $\delta^*(\boldsymbol{x})=11/102=0.108$.

接下来，假设 θ 的先验概率密度函数为 $\xi(\theta)=2(1-\theta)$，其中 $0<\theta<1$，用这种形式来代替均匀分布. 我们还是用样本容量为 100 的随机样本，其中 10 项是次品. 由于 $\delta(\theta)$ 是参数为 $\alpha=1$ 和 $\beta=2$ 的 β 分布的概率密度函数，从式(6.3.5)可以得出该情形下 θ 的贝叶斯估计是 $\delta(\boldsymbol{x})=11/103=0.107$.

这里所用的两个先验分布是有很大区别的. 以均匀分布为先验分布其均值是 1/2，而以 β 分布为先验分布其均值是 1/3. 尽管如此，由于样本中观察数是很大的，[350] 所以对于两种不同先验分布的贝叶斯估计几乎是相同的. 而且两种估计值都十分接近于样本中所观察到的次品的比例，即 $\bar{x}_n=0.1$.

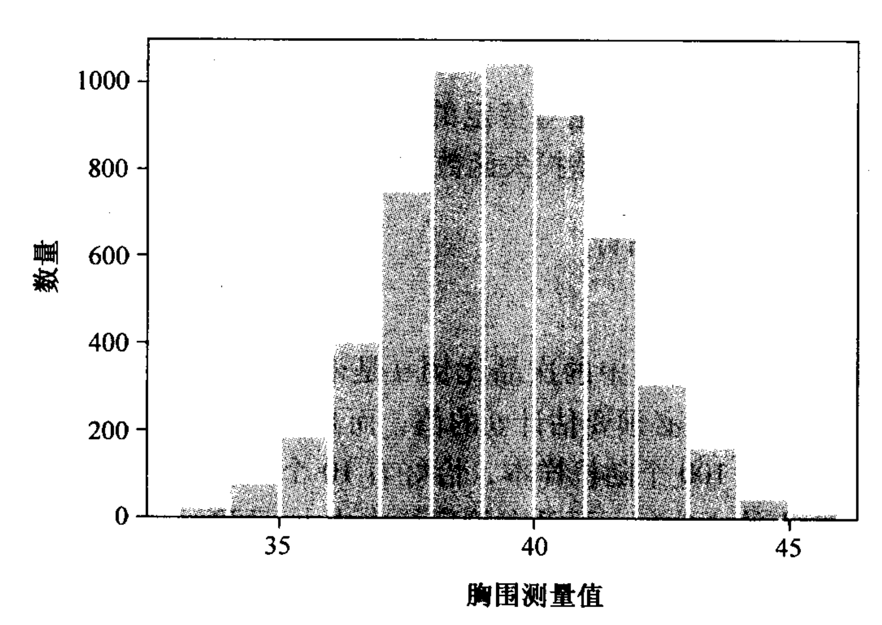
例 6.3.5 苏格兰士兵的胸围测量 Quetelet(1846)和一些记者报导了 5732 个苏格兰士兵测量胸围的数据. 这些数据最早出现在 1817 年的一个医学杂志上，并由 Stigler(1986)作了讨论. 图 6-1 显示了数据的直方图. 假设我们把单个胸围测量视为是来自均值是 θ 和方差是 4 的正态随机变量的随机样本(给定 θ). 胸围测量的平均值 $\bar{x}_n=39.85$. 如果 θ 有一个均值是 μ 且方差是 v^2 的先验正态分布，则由等式(3.6.1)，θ 的后验分布是正态分布，其均值为

$$\mu_1=\frac{4\mu+5732\times v^2\times 39.85}{4+5732\times v^2},$$

方差为

$$v_1^2=\frac{4v^2}{4+5732v^2}.$$

其贝叶斯估计值将会是 $\delta(\boldsymbol{x})=\mu_1$. 注意：除非 μ 惊人的大或者 v^2 非常地小，否则我们会有 μ_1 近似等于 39.85，v_1^2 近似等于 4/5732. 事实上，如果 θ 的先验概率密度函数是任意在 39.85 附近而不偏离很远的正的连续函数，那么 θ 的先验概率密度就十分接近于均值为 39.85 和方差为 4/5732 的正态概率密度函数. 不管是何种先验分布，其后验概率分布的均值和中位数都接近于 $\bar{x}_n$. ◀



351

图 6-1 例 6.3.5 中苏格兰士兵胸围测量的直方图

2. 贝叶斯估计的相容性

令 $X_1,\cdots,X_n$ 是一个来自于参数为 θ 的伯努利分布的随机样本. 假设我们对 θ 用其共轭先验分布. 由于 θ 是给定样本分布的均值，所有由 4.7 节中的大数定律可知当 $n\to\infty$ 时，$\overline{X}_n$ 依概率收敛于 θ. 既然贝叶斯估计量 $\delta^*(\boldsymbol{X})$ 与 $\overline{X}_n$ 的差在 $n\to\infty$ 时，依概率收敛于 0，那么可以推断当 $n\to\infty$ 时，$\delta^*(\boldsymbol{X})$ 依概率收敛于 θ 的未知值.

若 $n\to\infty$ 时，一个估计量的序列收敛于被估计参数的未知值，则称其为*估计的相容序列*. 因此在我们考虑的这个问题中，我们已经证明了贝叶斯估计 $\delta^*(\boldsymbol{X})$ 形成了一组估计量的相容序列. 该结论的实际意义就是：当得到大量的观察值时，贝叶斯估计量接近于未知参数 θ 的可能性会很高.

对于伯努利分布的参数估计中所得到的结论，也适用于其他估计问题. 在相当一般的情况下及对于广泛的损失函数族，某些参数 θ 的贝叶斯估计在样本容量 $n\to\infty$ 时，会形成估计量的相容序列. 特别地，对于任意来自于 6.3 节中讨论的各种分布族的随机样本，如果对于参数指定其共轭先验分布，并且使用平方损失函数，那么贝叶斯估计量就会组成一组估计量的相容序列.

例如，重新回到例 6.3.2 中的条件. 在该例子中，来自于均值 θ 为未知的正态分布的一个随机样本，其贝叶斯估计量 $\delta^*(\boldsymbol{X})$ 由式(6.3.6)给定. 依据大数定律，当 $n\to\infty$ 时，$\overline{X}_n$ 将收敛于未知均值 θ. 由式(6.3.6)可知，当 $n\to\infty$ 时，$\delta^*(\boldsymbol{X})$ 也收敛于

θ. 因此，贝叶斯估计量又形成了一组估计量的相容序列.

其他例子会在本节末的习题7和11中给出.

6.3.6 扩展研究：贝叶斯估计量的局限性

本节中所阐述的贝叶斯估计量的理论提供了一套令人满意且条理清晰的参数估计理论. 实际上，支持贝叶斯的统计学家们，认为它提供的只是条件清晰的估计理论. 然而，此套理论在实际统计问题的使用中还有一定的局限性. 为了应用，必须要先指定一个特别的损失函数，例如平均或绝对误差损失函数，并且还要为参数指定一个先验分布. 此类有意义的设定从理论上来说有可能是存在的，但要确定它们却可能很困难，而且很费时间. 在有些问题中，统计学家们必须决定适合于顾客或雇主的类型，而这些种类或者是得不到的，或者他们不会愿意交流他们的偏好和见解. 在其他一些问题中，一个估计有可能要由一群或是一个组织的成员共同努力来得到，但是通常要求组内的成员对于选用合适的损失函数和先验分布达成共识是相当困难的.

352

另外一个可能的难题是在某个特定的问题中，参数 θ 有可能是一个由所有值都未知的实值参数组成的向量. 由前面几章得到的贝叶斯估计的理论，可以很简单地被推广到含有向量参数 $\boldsymbol{\theta}$ 的估计的情形. 然而，要在此类问题中应用该理论，有必要先确定一个关于用来估计 $\boldsymbol{\theta}$ 所必需的多元先验分布，以及一个是向量 $\boldsymbol{\theta}$ 和向量 $\boldsymbol{a}$ 的函数的损失函数 $L(\boldsymbol{\theta},\boldsymbol{a})$. 即使统计人员可能只对估计问题中向量 $\boldsymbol{\theta}$ 的一两个分量感兴趣，他也必须要确定整个向量 $\boldsymbol{\theta}$ 的一个多元先验分布. 在很多重要的统计问题中，其中部分会在后面的章节中讨论到，$\boldsymbol{\theta}$ 可能有很多分量. 在此种问题中，要在多维参数空间 Ω 内确定一个有意义的先验分布，尤为困难.

应该强调的是没有简单的方法能够解决该类难题. 那些不是基于先验分布和损失函数的估计方法也明显地存在着应用的局限性，且在理论结构中也明显地存在着严重的缺陷.

6.3.7 小结

参数 θ 的估计量是数据 $\boldsymbol{X}$ 的一个函数 δ. 如果观察到 $\boldsymbol{X}=\boldsymbol{x}$，那么 $\delta(\boldsymbol{x})$ 的值就称为是我们的估计，即估计量 $\delta(\boldsymbol{X})$ 的观察值. 损失函数 $L(\theta,a)$ 是用来测量以 a 来估计 θ 时的损失. 选择贝叶斯估计量 $\delta^*(\boldsymbol{X})$，以使得当 $a=\delta^*(\boldsymbol{x})$ 函数 $L(\theta,a)$ 的后验均值达到最小值. 即：

$$\mathrm{E}[L(\theta,\delta^*(\boldsymbol{x}))|\boldsymbol{x}]=\min_a \mathrm{E}[L(\theta,a)|\boldsymbol{x}].$$

如果损失是平方误差 $L(\theta,a)=(\theta-a)^2$，那么 $\delta^*(\boldsymbol{x})$ 是 θ 的后验均值，即 $\mathrm{E}(\theta|\boldsymbol{x})$. 如果损失是绝对误差，$L(\theta,a)=|\theta-a|$，则 $\delta^*(\boldsymbol{x})$ 是 θ 后验分布的中位数. 对于其他损失函数，有可能要通过数值计算来确定其最小值.

习题

1. 在一个临床试验中，令成功治愈的概率 θ 的先验分布是[0,1]上的均匀分布. 假设第一个患者是一个成功治愈的病例. 在平方误差与绝对误差的损失函数下分别求 θ 的贝叶斯估计.
2. 假设在一批大宗运输货物中次品比例 θ 是未知的，θ 的先验分布是参数为 $\alpha=5$ 和 $\beta=10$ 的 β 分布. 假设从该批货物中随机抽取 20 件恰好有一件货物被发现是次品. 如果用平方误差损失函数，θ 的贝叶斯估计是什么?
3. 再次考虑习题 2 的情况. 假设 θ 的先验分布同习题 2，也同样假定从该批货物中随机抽取 20 件货物.

   a. 样本中有多少件次品时才能使贝叶斯估计的均方误差的均值最大?

   353

   b. 有多少件次品时会使贝叶斯估计的均方误差达到最小?
4. 假设一个容量为 n 的随机样本，来自于一个参数 θ 为未知的伯努利分布，θ 的先验分布是均值为 μ_0 的 β 分布. 证明 θ 的后验分布的均值是具有 $\gamma_n\overline{X}_n+(1-\gamma)\mu_0$ 形式的加权平均，并且证明当 $n\to\infty$ 时，$\gamma_n\to1$.
5. 假设长度为 1200 英尺的磁带中损伤处的数目服从泊松分布，其均值 θ 为未知，且 θ 的先验分布是一个参数 $\alpha=3$ 和 $\beta=1$ 的 Γ 分布. 随机抽取 5 卷磁带发现磁带有损伤处的数目为 2，2，6，0，3. 如果运用平方误差损失函数，那么 θ 的贝叶斯估计是多少?
6. 假设从均值 θ 未知的泊松分布中随机抽取一个容量为 n 的随机样本，θ 的先验分布是均值为 μ_0 的 Γ 分布. 证明：θ 的后验分布的均值是具有 $\gamma_n\overline{X}_n+(1-\gamma)\mu_0$ 形式的加权平均，且当 $n\to\infty$ 时，$\gamma_n\to1$.
7. 再次考虑习题 6 的条件，假设 θ 的值必须用平方误差损失函数来估计. 证明：贝叶斯估计量，$n=1,2,\cdots$，形成估计量 θ 的相容序列.
8. 假设某个总体中个体的身高服从均值 θ 未知、标准差为 2 英寸的正态分布. 假设 θ 的先验分布是一个均值为 68 英寸、标准差是 1 英寸的正态分布. 从该总体中随机抽取 10 个个体，他们的平均身高为 69.5 英寸.

   a. 如果用平方误差损失函数，θ 的贝叶斯估计是多少?

   b. 如果用绝对误差损失函数，θ 的贝叶斯估计是多少?
9. 假设从一个均值 θ 为未知、方差为 2 的正态分布中抽取一个随机样本，θ 的先验函数服从标准差为 1 的正态分布，θ 的值必须用平方误差损失函数来估计. 求使 θ 的贝叶斯估计量的均方误差为 0.01 甚至更小的最小样本容量为多少?
10. 假设一个顾客在某店接受服务所需时间服从参数 θ 值未知的指数分布，θ 的先验分布是均值为 0.2，标准差为 1 的 Γ 分布，且观察到一个容量为 20 的随机样本的平均服务时间是 3.8 分钟. 如果使用平方误差损失函数，θ 的贝叶斯估计是什么?
11. 假设一个容量为 n 的随机样本服从参数 θ 为未知的指数分布，θ 的先验分布是一个指定的 Γ 分布，而且 θ 的值必须用平方误差函数来估计. 证明：当 $n=1,2,\cdots$ 时，贝叶斯估计量形成 θ 的一个相容序列.
12. 令 θ 表示在大城市中支持某个提案的注册投票人的比例. 假设 θ 为未知，两个统计员 A 和 B 为 θ 分别指定如下不同的先验分布 $\xi_A(\theta)$ 和 $\xi_B(\theta)$：

$$\xi_A(\theta)=2\theta \quad 对于 0<\theta<1,$$
$$\xi_B(\theta)=4\theta^3 \quad 对于 0<\theta<1.$$

在由城市 1000 名注册投票人所组成的一个随机样本中，有 710 人支持这个提案.

a. 找出每个统计员给 θ 指定的后验分布.

b. 找出每个统计员在平方误差损失函数下的贝叶斯估计.

c. 证明在得到该随机样本中 1000 名注册投票人的意见后，不论样本中支持提案的人数是多少，两个统计员的贝叶斯估计相差不会超过 0.002.

13. 假设 $X_1,\cdots,X_n$ 是一个来自区间为 $[0,\theta]$ 上均匀分布的一个随机样本，其中参数 θ 为未知. 假设 θ 的先验分布是如 5.9 节中习题 16 中所定义的参数为 x_0 和 $\alpha(x_0>0,\alpha>0)$ 的帕累托分布. 如果用平方误差损失函数来估计 θ 的值，则 θ 的贝叶斯估计是什么？ 354

14. 假设 $X_1,\cdots,X_n$ 是来自于参数 θ 为未知的指数分布的一个随机样本，令为 $\xi(\theta)$ 表示 θ 的先验分布的概率密度函数，$\hat{\theta}$ 表示在先验概率密度函数为 $\xi(\theta)$ 的条件下使用平方误差损失函数时的贝叶斯估计. 令 $\psi=\theta^2$，并假设我们现在不估计 θ，而是希望在平方误差损失函数下估计 ψ 的值

$$L(\psi,a)=(\psi-a)^2, 当 \psi>0 及 a>0 时.$$

令 $\hat{\psi}$ 表示 ψ 的贝叶斯估计. 解释为什么 $\hat{\psi}>\hat{\theta}^2$. 提示：利用以下事实，对于有两个或两个以上取值的随机变量，有 $E(Z^2)>[E(Z)]^2$.

15. 令 $c>0$，考虑损失函数

$$L(\theta,a)=\begin{cases} c|\theta-a| & 对于 \theta<a, \\ |\theta-a| & 对于 \theta\geqslant a. \end{cases}$$

假设 θ 服从一个连续分布. 证明：θ 的贝叶斯估计将会是 θ 的后验分布的任何 $1/(1+c)$ 分位数. 提示：证明过程和定理 4.5.1 的过程很相似. 该结论即使当 θ 不服从连续分布时也成立，只不过证明过程要更繁琐一些.

6.4 极大似然估计

极大似然估计是避免使用先验分布和损失函数来选择参数估计量的方法，选择使其似然函数达到最大值的 θ 值为估计的值.

6.4.1 简介

在本节中，我们构造一个无需指定先验分布和损失函数就能建立一个估计量的简单方法. 称该方法为极大似然法，它是由 R. A. 费希尔在 1912 年提出的. 极大似然估计能应用于大多数问题，具有很强的直观性，且通常都能得到 θ 的一个合理的估计. 此外，如果样本很大，用该方法可以得到 θ 的一个非常好的估计. 因此，极大似然法有可能是统计中估计方法中最常用的一种.

6.4.2 极大似然估计的定义

假设随机变量 $X_1,\cdots,X_n$ 是一个来自于概率函数或概率密度函数为 $f(x|\theta)$ 的离散型分布或连续型分布的随机样本，其中参数 θ 属于某个参数空间 Ω. 其中 θ 可以是一个实值参数或是一个向量. 对于样本中每一个观察向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$，联合概率函数或联合概率密度函数的值通常用 $f_n(\boldsymbol{x}|\theta)$ 来表示. 与前相同，当把 $f_n(\boldsymbol{x}|\theta)$ 看作为给定向量 $\boldsymbol{x}$ 时 θ 的函数，则称其为*似然函数*.

假设此时，观察向量 $\boldsymbol{x}$ 来自于一个离散分布. 如果必须要选择 θ 的一个估计，那么我们不可能考虑 $\theta\in\Omega$ 中的任何值，因为不可能得到已经观察到的向量 $\boldsymbol{x}$ 的真值. 此外，假设当 θ 有一个特定的值，譬如，$\theta=\theta_0$ 时，得到实际观察向量 $\boldsymbol{x}$ 的概率 $f_n(\boldsymbol{x}|\theta)$ 非常大，而对其他 $\theta\in\Omega$ 的任何值都非常小. 自然地，我们会把 θ_0 当作 θ 的估计值（除非我们有超过样本中证据的很强的先验信息指向另外的值）. 当样本来自于一个连续型分布时，那么很自然要找到一个使得概率密度 $f_n(\boldsymbol{x}|\theta)$ 很大的 θ 值，并且用该值作为 θ 的估计值. 对于每个可能观察到的向量 $\boldsymbol{x}$，由于这个原因，我们会考虑使得似然函数 $f_n(\boldsymbol{x}|\theta)$ 达到最大值的 θ 值，并用该值作为 θ 的估计. 这个概念的严格定义如下：

355

对于每个可能观察到的向量 $\boldsymbol{x}$，令 $\delta(\boldsymbol{x})\in\Omega$ 表示 $\theta\in\Omega$ 中使得似然函数 $f_n(\boldsymbol{x}|\theta)$ 达到最大的值，并用 $\hat{\theta}=\delta(\boldsymbol{X})$ 来表示如此定义的 θ 的估计量. 称该估计量为 θ 的*极大似然估计*. *极大似然估计量或极大似然估计*可以简写为 M. L. E. .

6.4.3 极大似然估计的示例

值得指出的是在有些问题中，对于某些观察到的向量 $\boldsymbol{x}$，有可能对于 $\theta\in\Omega$ 中的任何点都不可能得到 $f_n(\boldsymbol{x}|\theta)$ 的极大值. 在此种情形中，θ 的 M. L. E. 不存在. 对于某些其他观察到的向量 $\boldsymbol{x}$，可能在空间 Ω 中不止一个点可以得到 $f_n(\boldsymbol{x}|\theta)$ 的极大值. 在此种情况下，M. L. E. 没有被唯一确定，而是这些点中任何一个值都可以被选作估计 $\hat{\theta}$. 尽管如此，在很多实际问题中，M. L. E. 是存在且可以唯一确定的 .

下面通过几个例子来说明极大似然方法以及各种可能性. 在每个例子中，我们都试图得到一个 M. L. E. .

例 6.4.1 一种疾病的检验 假设你沿着马路走时注意到公众健康部门正在给某种疾病做免费的医疗检验. 该测试从下面的角度来说其可信度为 90%：如果一个病人得了这种疾病，那么检验结果是阳性反应的概率为 0.9，然而，如果一个人没有得这种疾病，那么检验结果呈阳性反应的可能性仅为 0.1. 在例 2.3.1 中曾考虑过这一检验. 令 X 表示这一检验的结果，其中 $X=1$ 表示检验结果呈阳性，而 $X=0$ 表示检验结果为阴性. 令参数空间为 $\Omega=\{0.1,\ 0.9\}$，其中 $\theta=0.1$ 表示被检测病人没有得此种疾病，而 $\theta=0.9$ 表示此人已得此种疾病. 根据此参数空间，在给定 θ 的条件下，X 服从参数为 θ 的伯努利分布. 似然函数为

$$f(x|\theta)=\theta^x(1-\theta)^{1-x}.$$

如果观察到 $x=0$，那么

$$f(0|\theta)=\begin{cases}0.9 & \text{对于 } \theta=0.1,\\ 0.1 & \text{对于 } \theta=0.9.\end{cases}$$

356

非常清楚，当观察到的是 $x=0$，那么 $\theta=0.1$ 使得似然函数极大化. 如果观察的是 $x=1$，那么

$$f(1|\theta)=\begin{cases}0.1 & \text{对于 } \theta=0.1,\\ 0.9 & \text{对于 } \theta=0.9.\end{cases}$$

明显地，当观察到的是 $x=1$ 时，$\theta=0.9$ 使得似然函数极大化. 因此，我们有 M. L. E.

$$\hat{\theta}=\begin{cases}0.1 & \text{对于 } X=0,\\ 0.9 & \text{对于 } X=1.\end{cases}$$ ◀

例 6.4.2 来自伯努利分布的样本 假定随机变量 $X_1,\cdots,X_n$ 是来自于参数 θ 未知 $(0\leqslant\theta\leqslant1)$ 的伯努利分布的一个随机样本. 对于所有的观察值 $x_1,\cdots,x_n$，其中每个 x_i 的取值是 0 或 1，似然函数为

$$f_n(\boldsymbol{x}|\theta)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}. \tag{6.4.1}$$

由于对数函数是递增的，所以使似然函数 $f_n(\boldsymbol{x}|\theta)$ 极大化的 θ 值与使 $\ln f_n(\boldsymbol{x}|\theta)$ 极大化的 θ 值是相同的. 因而，通过找到使 $\ln f_n(\boldsymbol{x}|\theta)$ 达到极大的 θ 来决定 M. L. E. 要更为简便一些

$$\begin{aligned}L(\theta)&=\ln f_n(\boldsymbol{x}|\theta)=\sum_{i=1}^{n}[x_i\ln\theta+(1-x_i)\ln(1-\theta)]\\&=\left(\sum_{i=1}^{n}x_i\right)\ln\theta+\left(n-\sum_{i=1}^{n}x_i\right)\ln(1-\theta).\end{aligned} \tag{6.4.2}$$

计算导数 $\mathrm{d}L(\theta)/\mathrm{d}\theta$，并令导数等于 0，从而由结果方程中解得 θ. 如果 $\sum_{i=1}^{n}x_i\notin\{0,n\}$，我们发现当 $\theta=\bar{x}_n$ 时，导数为 0，而我们可以证明事实上这个值并没有使 $L(\theta)$ 和(6.4.1)式定义的式子达到极大值. 如果 $\sum_{i=1}^{n}x_i=0$，那么对于所有的 θ，$L(\theta)$ 是 θ 的一个递减函数，因此 L 在 $\theta=0$ 时达到其极大值. 类似地，如果 $\sum_{i=1}^{n}x_i=n$，L 是一个递增函数，那么它在 $\theta=1$ 时达到其极大值. 在最后两种情形中，注意到似然函数的极大值出现在 $\theta=\bar{x}_n$ 处. 从而可以得到 θ 的 M. L. E. 为 $\hat{\theta}=\bar{X}_n$. ◀

从例 6.4.2 可以看到如果把 $X_1,\cdots,X_n$ 看作 n 重伯努利试验，那么在任意给定试验中未知的成功概率的 M. L. E. 就是 n 个试验中观察到的成功的比例.

357

例6.4.3 来自正态分布的样本 假设 $X_1,\cdots,X_n$ 是来自于均值 μ 未知、σ^2 已知的正态分布的一个随机样本. 对于所有观察值 $x_1,\cdots,x_n$,似然函数 $f_n(\boldsymbol{x}|\mu)$ 为

$$f_n(\boldsymbol{x}|\mu) = \frac{1}{2(\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]. \tag{6.4.3}$$

从式(6.4.3)中可以看出使下式

$$Q(\mu) = \sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}x_i^2 - 2\mu\sum_{i=1}^{n}x_i + n\mu^2.$$

极小化的 μ 的值会使 $f_n(\boldsymbol{x}|\mu)$ 达到极大.

如果现在计算导数 $\mathrm{d}Q(\mu)/\mathrm{d}\mu$, 令其为0, 然后从得到的等式中求出 μ,发现 $\mu=\bar{x}_n$. 从而, μ 的 M. L. E. 为 $\hat{\mu}=\bar{X}_n$. ◀

从例6.4.3可见, 估计量 $\hat{\mu}$ 不受方差 σ^2 的值的影响, 其中假定 σ^2 已知. 不管 σ^2 为何值, 未知均值 σ 的 M. L. E. 总是简单的样本均值 $\bar{X}_n$. 在下面的例子中再来看看 μ 和 σ^2 的估计.

例6.4.4 从未知方差的正态分布中抽样 假设 $X_1,\cdots,X_n$ 是来自于正态分布的一个随机样本, 但假设 μ 和 σ^2 都未知. 则参数 $\theta=(\mu,\sigma^2)$. 对于所有观察值 $x_1,\cdots,x_n$ 似然函数 $f_n(\boldsymbol{x}|\mu,\sigma^2)$ 就如等式(6.4.3)式中的右半部分. 该函数必须在 μ 和 σ^2 所有可能取值中极大化, 其中 $-\infty<\mu<\infty$ 且 $\sigma^2>0$. 不直接使似然函数 $f_n(\boldsymbol{x}|\mu,\sigma^2)$ 直接极大化, 而是使 $\ln f_n(\boldsymbol{x}|\mu,\sigma^2)$ 极大化更为简单. 我们有:

$$\begin{aligned} L(\theta) &= \ln f_n(\boldsymbol{x}|\mu,\sigma^2) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \end{aligned} \tag{6.4.4}$$

我们应该分三个步骤来找到使 $L(\theta)$ 极大化的 $\theta=(\mu,\sigma^2)$ 的值. 首先, 对于每个固定的 σ^2, 我们要找到使(6.4.4)式右边达到极大的 $\hat{\mu}(\sigma^2)$ 的值. 其次, 当 $\theta'=(\hat{\mu}(\sigma^2),\sigma^2)$ 时, 找到使 $L(\theta')$ 达到极大的值 σ^2. 最后, θ 的极大似然估计是随机向量. 它的观察值是 $(\hat{\mu}(\hat{\sigma}^2),\hat{\sigma}^2)$. 第一步在例6.4.3中已经解决. 当时已经得到 $\hat{\mu}(\sigma^2)=\bar{x}_n$. 对于第二步, 令 $\theta'=(\bar{x}_n,\sigma^2)$ 并使下式极大化

$$L(\theta') = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2. \tag{6.4.5}$$

可以通过对上式中的 σ^2 求导数, 并使其等于0, 从而解出 σ^2 使上式最大化. 导数为

358

$$\frac{\mathrm{d}}{\mathrm{d}\sigma^2}L(\theta') = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2.$$

令其为0, 可以得到

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2. \tag{6.4.6}$$

(6.4.5)式在 σ^2 取(6.4.6)式中的值时二次导数为负，所以我们已经找到了最大值. 因此，$\theta=(\mu,\sigma^2)$的 M. L. E. 为

$$\hat{\theta}=(\hat{\mu},\hat{\sigma}^2)=\left(\overline{X}_n,\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right). \quad (6.4.7)$$

回顾式(6.4.7)极大似然估计中的第一个坐标被称为数据的样本均值. 同理，我们称 M. L. E. 中第二个坐标为样本方差. 不难发现，样本方差的观察值是给样本 n 个观察值 $x_1,\cdots,x_n$ 中每个值都指定为 $1/n$ 的概率的分布的方差（见习题 1）.

例 6.4.5 来自均匀分布的样本 假设 $X_1,\cdots,X_n$ 是来自于区间$[0,\theta]$上均匀分布的一个随机样本，其中参数 θ 未知($\theta>0$)，每个观察值的概率密度 $f(x|\theta)$都具有如下形式：

$$f(x|\theta)=\begin{cases}\frac{1}{\theta} & 对于 0\leqslant x\leqslant\theta,\\ 0 & 其他.\end{cases} \quad (6.4.8)$$

因此，$X_1,\cdots,X_n$ 的联合概率密度函数 $f_n(\boldsymbol{x}|\theta)$为

$$f_n(\boldsymbol{x}|\theta)=\begin{cases}\frac{1}{\theta^n} & 对于 0\leqslant x_i\leqslant\theta(i=1,\cdots,n),\\ 0 & 其他.\end{cases} \quad (6.4.9)$$

从式(6.4.9)可以看到，θ 的 M. L. E. 一定是满足 $\theta\geqslant x_i,i=1,\cdots,n$ 且使得$1/\theta^n$ 在所有值中极大化的 θ 值. 由于 $1/\theta^n$ 是 θ 的一个递减函数，估计将会是满足$\theta\geqslant x_i$，$i=1,\cdots,n$ 中 θ 的最小值. 既然这个值是 $\theta=\max(x_1,\cdots,x_n)$，那么 θ 的 M. L. E. 就是 $\hat{\theta}=\max(X_1,\cdots,X_n)$.

6.4.4 扩展研究：极大似然估计的局限性

尽管直观上可以理解，但极大似然估计并不一定适用于所有的问题. 例如，在例 6.4.5 中，M. L. E. $\hat{\theta}$ 看上去就不是 θ 的一个合适的估计量. 由于 $\max(X_1,\cdots,X_n)<\theta$ 的概率为 1，所以可以肯定 $\hat{\theta}$ 低估了 θ 的值. 事实上，如果已知 θ 的任意先验分布，那么 θ 的贝叶斯估计量无疑要比 $\hat{\theta}$ 大. 当然贝叶斯估计量实际超出 $\hat{\theta}$ 的量，要取决于所用的特定的先验分布和观察到的 $X_1,\cdots,X_n$ 的值. 例 6.4.5 也指出了极大似然的另一个困难，我们将在例 6.4.6 中说明.

例 6.4.6 M. L. E. 不存在 仍设 $X_1,\cdots,X_n$ 是来自于区间$[0,\theta]$上均匀分布的一个随机样本. 不过这次不用(6.4.8)式所给定的均匀分布概率密度函数 $f(x|\theta)$ 的形式，而是改用以下形式：

$$f(x|\theta)=\begin{cases}\dfrac{1}{\theta} & 对于 0<x<\theta,\\ 0 & 其他.\end{cases} \tag{6.4.10}$$

(6.4.8)式与(6.4.10)式的唯一区别在于(6.4.10)式中的严格不等式代替了(6.4.8)式中的弱不等式，从而改变了概率密度函数在两端点0和θ的值. 因而，任意一个式子都可以被当作均匀分布的密度函数. 然而如果把(6.4.10)式作为概率密度函数，那么θ的M. L. E. 是满足$\theta>x_i$，$i=1,\cdots,n$，且在所有值中使$1/\theta^n$达到最大的值. 值得一提的是在θ所有可能取值中已经不包括$\theta=\max(x_1,\cdots,x_n)$，因为$\theta$必须严格大于每个观察值$x_i$，因为$\theta$可以取与$\max(x_1,\cdots,x_n)$任意接近的值但却不可以等于这个值，这就导致了M. L. E. 不存在. ◀

在所有我们以前对于概率密度函数的讨论中，强调了以下的事实：概率密度函数为$1/\theta$的均匀分布取值在开区间$0<x<\theta$还是在闭区间$0\leqslant x\leqslant\theta$是无关紧要的. 但现在看到M. L. E. 的存在性就取决于该不相干和不重要的选择. 只要使用(6.4.8)式中所定义的密度函数，而不使用(6.4.10)式中所定义的，就可以非常容易避免这一困难. 在许多其他问题中，可以通过选择一个已有概率密度函数特定的合适的形式来代替已给定的分布，从而避免此类困难. 然而，在例6.4.8中将看到，并不是总能避免此类困难的.

例6.4.7 不唯一的M. L. E. 设$X_1,\cdots,X_n$是取自区间$[\theta,\theta+1]$上均匀分布的一个随机样本，其中参数θ未知($-\infty<\theta<\infty$). 在本例中，联合概率密度函数有以下形式

$$f_n(\boldsymbol{x}|\theta)=\begin{cases}1 & 对于 \theta\leqslant x_i\leqslant\theta+1(i=1,\cdots,n),\\ 0 & 其他.\end{cases} \tag{6.4.11}$$

条件$\theta\leqslant x_i$，$i=1,\cdots,n$等价于条件$\theta\leqslant\min(x_1,\cdots,x_n)$. 类似地，条件$x_i\leqslant\theta+1$等价于条件$\theta\geqslant\max(x_1,\cdots,x_n)-1$. 这样，我们不把$f_n(\boldsymbol{x}|\theta)$表示成(6.4.11)式中所给的形式，而是使用如下形式

$$f_n(\boldsymbol{x}|\theta)=\begin{cases}1 & \max(x_1,\cdots,x_n)-1\leqslant\theta\leqslant\min(x_1,\cdots,x_n),\\ 0 & 其他.\end{cases} \tag{6.4.12}$$

这样可以在

$$\max(x_1,\cdots,x_n)-1\leqslant\theta\leqslant\min(x_1,\cdots,x_n). \tag{6.4.13}$$

中选择任意θ值作为M. L. E. .

360

在本例中，$\hat{\theta}$不是唯一确定的. 实际上，极大似然法对于选择θ的一个估计提供了很小的帮助. 对于每一个不在(6.4.13)区间的θ的值的似然估计都是θ，因此在该区间以外的θ值不会被估计，而且在区间之内的所有值都是M. L. E. . ◀

例 6.4.8 来自两种混合分布的样本 考虑随机变量 X 可以等可能地来自于均值为 0、方差为 1 的正态分布，或是来自于另一个均值为 μ、方差为 σ^2 的正态分布，其中 μ 和 σ^2 都未知. 在该条件下，X 的概率密度函数将是这两个不同正态分布概率密度函数的平均. 于是，

$$f(x|\mu,\sigma^2)=\frac{1}{2}\left\{\frac{1}{(2\pi)^{1/2}}\exp\left(-\frac{x^2}{2}\right)+\frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\right\}. \quad (6.4.14)$$

设随机样本来自于概率密度函数如(6.4.14)式中所定义的分布. 如前所示，似然函数 $f_n(\boldsymbol{x}|\mu,\sigma^2)$ 具有以下形式

$$f_n(\boldsymbol{x}|\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i|\mu,\sigma^2). \quad (6.4.15)$$

为了找到 μ，σ^2 的 M. L. E.，必须找到使得 $f_n(\boldsymbol{x}|\mu,\sigma^2)$ 达到最大的 μ 和 σ^2 的值.

用 x_k 表示 $x_1,\cdots,x_n$ 中任意一个观察值. 如果令 $\mu=x_k$，且让 $\sigma^2\to 0$，则在(6.4.15)式右端的因子 $f(x_k|\mu,\sigma^2)$ 会变成无穷大，而对于每个 $x_i\neq x_k$，因子 $f(x_i|\mu,\sigma^2)$ 都会逼近于值

$$\frac{1}{2(2\pi)^{1/2}}\exp\left(-\frac{x_i^2}{2}\right).$$

因此当 $\mu=x_k$ 和 $\sigma^2\to 0$ 时，我们发现 $f_n(\boldsymbol{x}|\mu,\sigma^2)\to\infty$.

0 不是 σ^2 的一个允许估计值，因为我们已经预知 $\sigma^2>0$. 由于当选择 $\mu=x_k$ 以及任意小接近于 0 的 σ^2 时，似然函数可以达到任意大，这意味着 M. L. E. 不存在.

如果我们允许 0 为 σ^2 的一个允许估计来克服此困难，那么发现 μ 和 σ^2 有 n 对不同的 M. L. E.，即 $\hat{\mu}=x_k$ 和 $\hat{\sigma}^2=0$，当 $k=1,\cdots,n$. 这些估计值看起来是荒唐的. 再回想例子开始的描述，观察值可能来自于两个正态分布中的哪个分布. 比如，设 $n=1000$，我们用估计 $\hat{\mu}=x_3$，$\hat{\sigma}^2=0$. 那么我们会估计未知方差的值为 0，还有，更有效地，我们可以推断 x_3 是来自于给定的未知分布，而其他 999 个观察值来自于那个标准正态分布. 事实上，既然每个观察值等可能地来自于这两个分布，那么应该有数百个观察值，而不是一个来自于未知的正态分布. 在本例中，极大似然方法明显地不适用. 在 11.3 节的习题 6 中将提及该问题的贝叶斯解法. ◀ 361

最后，我们将提到关于 M. L. E. 的一个说明. M. L. E. 是使数据在给定 θ 条件下条件概率函数或概率密度函数极大化的 θ 值. 因此，极大似然估计是产生观察数据最可能的 θ 值. 参数的值未必看上去是与给出数据最接近的那个. 若求参数取不同值的可能性有多大，就需要该参数的概率分布. 当然，该参数的后验分布(6.2 节)将用于此目的，但是 M. L. E. 的计算中没有涉及到后验分布. 因此，在观测到数据之后将极大似然估计作为参数最可能的值是不恰当的. 举个例子，考虑例 6.4.2 中的情

形，假设我们将一枚硬币投掷几次，所关心的是它是否轻微地偏向于正面或反面. 令 $X_i=1$，如果第 i 次投掷是正面，否则令 $X_i=0$. 如果在前5次投掷中，我们得到4次正面和1次反面，观察到的 M. L. E. 的值将是0.8，但是很难想像在这种情况下，仅基于5次投掷而认为出现正面的概率，即 θ 的最可能取值会大到0.8，而这是发生在一枚以前看上去是均匀的硬币上. 将极大似然估计量视为一个参数的最可能的取值与例2.3.1和2.3.6中医学实验中忽视罕见疾病的先验信息是相似的. 如果此类例子中实验结果是阳性的，我们发现(在例6.4.1中)M. L. E. 的取值为 $\hat{\theta}=0.9$，与得此疾病相对应. 然而，如果得病的先验概率像例2.3.1中的那样小，试验结果是阳性，得病的后验概率依然非常小($\theta=0.9$). 实验无法精确到足以完全克服先验信息. 对于我们投掷硬币也一样，5次投掷没有足够的信息去克服硬币是一枚均匀硬币的先验信念. 只有当数据比可用的先验信息能获得更多信息时，才能近似正确地认为参数最有可能的取值就在 M. L. E. 值的附近. 这可能发生在当 M. L. E. 是建立在大量数据上或只有很少的先验信息的情况下.

6.4.5 小结

参数 θ 的极大似然估计是似然方程 $f_n(\boldsymbol{x}|\theta)$ 在给定数据 $\boldsymbol{x}$ 条件下提供最大值的那个 θ 值. 如果设 $\delta(\boldsymbol{x})$ 为极大似然估计，那么 $\hat{\theta}=\delta(\boldsymbol{X})$ 是极大似然估计量(M. L. E.). 我们已经计算过，当数据是来自于伯努利分布，方差未知的正态分布，两个参数都未知的正态分布，或在区间 $[0,\theta]$ 上均匀分布的一个随机样本时的 M. L. E. .

习题

1. 令 $x_1,\cdots,x_n$ 是不同的数. 令 Y 是一个离散的随机变量，其概率函数如下：

$$f(y)=\begin{cases}\dfrac{1}{n} & \text{对于 } y\in\{x_1,\cdots,x_n\},\\ 0 & \text{其他.}\end{cases}$$

证明：Y 的方差 $\mathrm{Var}(Y)$ 有如(6.4.6)式中给出的形式.

2. 购买某种商标的谷物早餐时由女性来决定的比例 p 和由男性来决定的比例未知. 在一个有70个购买者的随机样本中，发现其中58位是女性，12位是男性. 求 p 的极大似然估计量.

362

3. 再次考虑习题2中的情况，但是假设已知 $\frac{1}{2}\leqslant p\leqslant\frac{2}{3}$. 如果对于70个购买者的随机样本的观察情况和习题2中的一样，p 的 M. L. E. 是什么？

4. 假设 $X_1,\cdots,X_n$ 来自于一个参数 θ 未知的伯努利分布的随机样本，但是知道 θ 落在开区间 $0<\theta<1$ 中. 试证：如果每一个观察值都是0或每一个观察值都是1时，θ 的 M. L. E. 不存在.

5. 假设 $X_1,\cdots,X_n$ 是一个来自于均值 θ 未知的泊松分布的随机样本($\theta>0$).

   a. 假定至少有一个观察值不为0，试确定 θ 的 M. L. E. .

b. 如果每一个观察值都为 0，试证：θ 的 M. L. E. 不存在.

6. 假设 $X_1,\cdots,X_n$ 是一个来自于均值 μ 已知且方差 σ^2 未知的正态分布的随机样本. 求 σ^2 的 M. L. E..

7. 假设 $X_1,\cdots,X_n$ 是一个来自于参数 β 的值未知的指数分布的随机样本($\beta>0$)，求 β 的 M. L. E..

8. 假设 $X_1,\cdots,X_n$ 是一个来自于密度函数为 $f(x|\theta)$ 的分布的随机样本.

$$f(x|\theta)=\begin{cases}e^{\theta-x} & \text{对于 } x>\theta,\\ 0 & \text{对于 } x\leqslant\theta.\end{cases}$$

同样假设 θ 的值未知($-\infty<\theta<\infty$).

a. 试证：θ 的 M. L. E. 不存在.

b. 定义同一分布的概率密度函数的另一种形式，使得 θ 的 M. L. E. 存在，并求出它的估计量.

9. 假设 $X_1,\cdots,X_n$ 是一个来自于概率密度函数为 $f(x|\theta)$ 的分布的随机样本.

$$f(x|\theta)=\begin{cases}\theta x^{\theta-1} & \text{对于 } 0<x<1,\\ 0 & \text{其他.}\end{cases}$$

同时，假设 θ 的值未知($\theta>0$)，求 θ 的 M. L. E..

10. 假设 $X_1,\cdots,X_n$ 是一个来自于概率密度函数为 $f(x|\theta)$ 的分布的随机样本.

$$f(x|\theta)=\frac{1}{2}e^{-|x-\theta|}\text{对于 }-\infty<x<\infty.$$

同样，假设 θ 的值未知($-\infty<\theta<\infty$)，求 θ 的 M. L. E.. 提示：当分布如习题 1 中所定义时与使 M. A. E. 极小化问题作比较.

11. 假设 $X_1,\cdots,X_n$ 是一个区间 $[\theta_1,\theta_2]$ 上均匀分布的随机样本，其中 θ_1 和 θ_2 都未知($-\infty<\theta_1<\theta_2<\infty$). 求 θ_1 和 θ_2 的 M. L. E..

12. 假设某个很大的总体包含 k 种不同的个体($k\geqslant2$)，令 θ_i 表示第 i 种类型的比例，$i=1,\cdots,k$. 这里 $0\leqslant\theta_i\leqslant1$，且 $\theta_1+\cdots+\theta_k=1$. 同样，假设由 n 个个体组成的随机样本来自于该总体，且恰好有 n_i 个个体在类型 i 中，其中 $n_1+\cdots+n_k=n$. 求 $\theta_1,\cdots,\theta_k$ 的 M. L. E..

13. 假设二维向量 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 是一个来自于二维正态分布的随机样本，其中 X 和 Y 的均值未知但 X 和 Y 的方差以及 X 和 Y 的相关系数都是已知的. 求均值的 M. L. E..

14. 假设二维向量 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 是一个来自于二维正态分布的随机样本，其中 X 和 Y 的均值，X 和 Y 的方差以及 X 和 Y 的相关系数都未知. 证明：这 5 个参数的 M. L. E. 如下：

$$\hat{\mu}_1=\overline{X}_n,\hat{\mu}_2=\overline{Y}_n,$$

$$\hat{\sigma}_1^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2,\hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y}_n)^2,$$

$$\hat{\rho}=\frac{\sum_{i=1}^{n}(X_i-\overline{X}_n)(Y_i-\overline{Y}_n)}{\left[\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right]^{1/2}\left[\sum_{i=1}^{n}(Y_i-\overline{Y}_n)^2\right]^{1/2}}.$$

6.5 极大似然估计的性质

在本节中，我们要探讨 M. L. E. 的一些性质，包括

- 一个参数的 M. L. E. 和该参数的函数的 M. L. E. 的关系.
- 对计算算法的需求.
- 当样本容量变大时 M. L. E. 的变化.
- 在抽样方案中 M. L. E. 相关性的缺失.

6.5.1 不变性

假定随机变量 $X_1,\cdots,X_n$ 是来自于概率函数或概率密度函数为 $f(x|\theta)$ 的分布的一个随机样本，其中参数 θ 的值是未知的. 该参数可能是一维的或参数向量. 现在用 $\hat{\theta}$ 来表示参数 θ 的 M. L. E.. 因此，对于所有的观察值 $x_1,\cdots x_n$，似然函数 $f_n(\boldsymbol{x}|\theta)$ 在 $\theta=\hat{\theta}$时达到最大.

如果我们对于以上分布中的参数 θ 做如下的变换：不将概率函数或概率密度函数 $f(x|\theta)$ 表示为参数 θ 的形式，而是表示为新参数 $\tau=g(\theta)$ 的形式，其中 g 为 θ 的一一对应的函数. 新的参数空间 Γ 等于 Ω 在函数 g 下的像. 我们再令 $\theta=h(\tau)$ 表示 τ 的反函数. 因此，每个观察到的值的概率函数或概率密度函数用新的参数 τ 表示就可以写成 $f[x|h(\tau)]$，它的似然函数也变为 $f_n[\boldsymbol{x}|h(\tau)]$.

在 $f_n[\boldsymbol{x}|h(\tau)]$ 取得最大值时，参数 τ 的 M. L. E. $\hat{\tau}$ 就与该最大值对应. 因为 $f_n(\boldsymbol{x}|\theta)$ 在 $\theta=\hat{\theta}$ 时取得最大值. 因此 $f_n[\boldsymbol{x}|h(\tau)]$ 在 $h(\tau)=\hat{\theta}$ 时取得最大值. 因此，M. L. E. $\hat{\tau}$ 必须满足关系式 $h(\hat{\tau})=\hat{\theta}$，或者等价地 $\hat{\tau}=g(\hat{\theta})$. 因此我们就推导出了极大似然估计的下面这条性质，称为极大似然估计的不变性：

如果 $\hat{\theta}$ 是 θ 的极大似然估计，且函数 g 为一一对应的函数，那么 $g(\hat{\theta})$ 也是 $g(\theta)$ 的极大似然估计.

不变性的性质可以推广到非一一对应的函数情形. 例如，如果在均值和方差都未知的情况下，要估计正态分布的均值 μ，则 μ 就不是参数 $\theta=(\mu,\sigma^2)$ 一一对应的函数. 在此情况下，我们想要估计的是 $g(\theta)=\mu$. 有一种方法可以得到不必是一一对应的函数 θ 的 M. L. E. 通常的一种方法如下：令 $g(\theta)$ 表示 θ 的任意函数，令 G 是 Ω 在函数 g 下的象. 对于任一 $t\in G$，定义 $G_t=\{\theta|g(\theta)=t\}$和

$$L^*(t)=\max_{\theta\in G_t}\ln f_n(\boldsymbol{x}|\theta).$$

最后，定义 $g(\theta)$ 的 M. L. E. 为 $\hat{\tau}$，其中 $\hat{\tau}$ 满足

364

$$L^*(\hat{t})=\max_{t\in G}L^*(t). \tag{6.5.1}$$

定理 6.5.1 令 $\hat{\theta}$ 为 θ 的 M. L. E. $g(\theta)$ 为参数 θ 的函数. 那么 $g(\hat{\theta})$ 是 $g(\theta)$ 的 M. L. E..

证明：我们可以证明 $\hat{\tau}=g(\hat{\theta})$ 满足(6.5.1)．由于 $L^*(t)$ 是对于所有 θ 在 Ω 子集中 $\ln f_n(\boldsymbol{x}|\theta)$ 的最大值，而且 $\ln f_n(\boldsymbol{x}|\hat{\theta})$ 对于所有 θ 是最大的，由此我们可以推出对于所有的 $t\in G$，有 $L^*(t)\leqslant\ln f_n(\boldsymbol{x}|\hat{\theta})$．令 $t=g(\hat{\theta})$，接下来是证明关系式 $L^*(\hat{\tau})=\ln f_n(\boldsymbol{x}|\hat{\theta})$．注意到 $\hat{\theta}\in G_{\hat{\tau}}$，因为 $\hat{\theta}$ 是在所有的 θ 中使 $f_n(\boldsymbol{x}|\theta)$ 最大化，那么对于所有 $\theta\in G_{\hat{\tau}}$ 时，它也使 $f_n(\boldsymbol{x}|\theta)$ 达到最大．因此，$L^*(\hat{\tau})=\ln f_n(\boldsymbol{x}|\hat{\theta})$ 和 $\hat{\tau}=g(\hat{\theta})$ 是 $g(\theta)$ 的一个 M. L. E.． ■

例 6.5.1 标准差和二阶矩的估计 假定变量 $X_1,\cdots,X_n$ 是来自于均值 μ 和方差 σ^2 均未知的正态分布的一个随机样本．我们要确定标准差 σ 的 M. L. E. 和正态分布的二阶矩 $\mathrm{E}(X^2)$ 的 M. L. E. 在例 6.4.4 中已经给出了 $\theta=(\mu,\sigma^2)$ 的 M. L. E. 是 $\hat{\theta}=(\hat{\mu},\hat{\sigma}^2)$．由不变性可知，标准差的 M. L. E. 就是样本方差的平方根．用符号表示就是 $\sigma^2=(\hat{\sigma}^2)^{1/2}$．那么，因为 $\mathrm{E}(X^2)=\sigma^2+\mu^2$，$\mathrm{E}(X^2)$ 的 M. L. E. 就是 $\hat{\sigma}^2+\hat{\mu}^2$． ◀

6.5.2 数值计算

在许多问题中，对于一个给定的参数 θ 存在唯一的一个 M. L. E. $\hat{\theta}$，但是该 M. L. E. 不能用一个样本观察值的显式代数函数来表示．因此对于一个给定的观察值集合，就必须用数值计算来决定 $\hat{\theta}$ 的值．下面我们就用两个例子来对此做进一步的说明．

例 6.5.2 来自 Γ 分布的样本 设随机变量 $X_1,\cdots X_n$ 是来自于 Γ 分布的一个随机样本，其密度函数如下：

$$f(x|\alpha)=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-x}\qquad 对于\ x>0. \tag{6.5.2}$$

假设 α 的值也未知($\alpha>0$)且要对 α 做估计．

似然函数为

$$f_n(\boldsymbol{x}|\alpha)=\frac{1}{\Gamma^n(\alpha)}\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\exp\left(-\sum_{i=1}^n x_i\right). \tag{6.5.3}$$

α 的 M. L. E. 为满足下式的 α 的值

$$\frac{\partial\ln f_n(\boldsymbol{x}|\alpha)}{\partial\alpha}=0. \tag{6.5.4}$$

对于该例子使用式(6.5.4)，

就得到下式：

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}=\frac{1}{n}\sum_{i=1}^n\ln x_i. \tag{6.5.5}$$

365

在许多不同的已发表的数学表总汇中都可以查到被称为双伽玛函数 $\Gamma'(\alpha)/\Gamma(\alpha)$ 的表格．在一些数学软件中也可以得到该双伽玛函数．对于 $x_1,\cdots,x_n$ 所有给定的值，满足式(6.5.5) α 的唯一值，必须参照这些数学表或者是用双伽玛函数的数值计算来得到．该值就是 α 的 M. L. E.． ◀

例6.5.3　来自于柯西分布的样本　假设随机变量 $X_1,\cdots X_n$ 是来自某个未知点 $\theta(-\infty<\theta<\infty)$ 为中心的柯西分布的一个随机样本，其密度函数如下：

$$f(x|\theta)=\frac{1}{\pi[1+(x-\theta)^2]}\quad \text{对于} -\infty<x<\infty. \tag{6.5.6}$$

同时假设 θ 的值是待估计的.

似然函数是

$$f_n(\boldsymbol{x}|\theta)=\frac{1}{\pi^n\prod_{i=1}^n[1+(x_i-\theta)^2]}. \tag{6.5.7}$$

因此，θ 的 M. L. E. 将会是使下式达到最小的值

$$\prod_{i=1}^n[1+(x_i-\theta)^2]. \tag{6.5.8}$$

对于大部分的 $x_1,\cdots,x_n$ 使表达式(6.5.8)最小化的 θ 值必须通过一个数值计算来决定. ◀

6.5.3　相容性

考虑一个估计问题，该问题中的随机变量来自于一个包含参数 θ 的分布. 假设对于每一个充分大的样本容量 n，即，对于大于某个给定的最小数的 n 的每个值，都存在唯一一个 M. L. E. 那么，在某些条件下，尤其是在实际问题中满足的这些条件，M. L. E. 序列是 θ 的估计量的相容序列. 即在大部分问题中 M. L. E. 序列当 $n\to\infty$ 时依概率收敛于未知量 θ.

我们已经在6.3节中指出在某些一般条件下，参数 θ 的贝叶斯估计序列也是估计的一个相容序列. 因此，对于一个给定的先验分布和一个充分大的样本容量 n，θ 的贝叶斯估计和 M. L. E. 会相互接近，而且两者都会十分接近于 θ 的未知值.

我们不会给出证明该结果而需要的条件的任何细节(具体可以看参考文献 Schervish1995 的第7章). 然而我们会通过考虑来自于参数 θ 未知（$0\leqslant\theta\leqslant1$）的伯努利分布的一个随机样本 $X_1,\cdots,X_n$ 来说明该结果. 在6.3节中已经证明了如果给出的 θ 的先验分布是一个 β 分布，那么 θ 的贝叶斯估计和样本均值 $\overline{X}_n$ 的差当 $n\to\infty$ 时收敛于0. 此外，例6.4.2已给出了 θ 的 M. L. E. 是 $\overline{X}_n$. 因此，当 $n\to\infty$ 时，贝叶斯估计和 M. L. E. 的差将收敛于0. 最后，如6.3节中所提到的，样本均值 $\overline{X}_n$ 当 $n\to\infty$ 时，依概率收敛于 θ. 所以，贝叶斯估计序列和 M. L. E. 序列都是相容序列.

366

6.5.4　扩展研究：抽样方案

假设一位试验者想要从一个概率函数或密度函数是 $f(x|\theta)$ 的分布取得观察值以获取关于参数 θ 值的信息. 试验者可以简单地按照事先确定的样本容量从分布中抽取一个随机样本. 然而若不是这样，也许首先他会从分布中随机观察一些值，并注意得到这些观察值所需要花费的成本和时间. 然后他会确定再从该分布中随机观察

更多的一些值，并研究所有已得到的值. 试验者会决定停止收集观察值，并用所有已经得到的观察值去估计 θ 的值. 他决定停止的原因或许是因为他感到已经有了足够的信息能够得到 θ 的一个好的估计，也有可能是他没有能力去花费更多的钱或时间去取样了.

在该实验中，观察的样本数量 n 事先没有固定. 它是个随机变量且会相当依赖于他们所得的观察值的数量.

不管试验者是否决定在收集任何观察值前固定 n 的值或者喜欢用其他的抽样方案，例如刚才所描述的那处，都可以证明基于观察值的似然函数 $L(\theta)$ 可以写成

$$L(\theta)=f(x_1|\theta)\cdots f(x_n|\theta).$$

因此，无论使用何种形式的抽样方案，θ 的 M. L. E. 都将是相同的. 即 θ 的值由实际观察到的 $x_1,\cdots,x_n$ 值确定，而不取决于试验者用来决定什么时候停止抽样而所使用的方案(如果有的话).

为了说明该性质，假设在一个服务社区里的顾客相继到达的时间间隔用分钟表示是独立同分布的随机变量. 也假设每个时间间隔服从参数为 β 的指数分布，并且观察到的时间间隔 $X_1,\cdots X_n$ 的集合也是来自于该分布的一个随机样本. 从6.4节习题7可知，M. L. E. 将会是 $\hat{\beta}=1/\overline{X}_n$. 同样，既然指数分布的均值 μ 是 $1/\beta$，由 M. L. E. 的不变性性质可知：$\hat{\mu}=\overline{X}_n$. 即均值的 M. L. E. 是样本观察值的平均数.

现在考虑如下三个抽样方案：

(1) 一位试验者事先决定恰好抽取20个观察值，而且发现这20个数的平均数是6，那么 μ 的 M. L. E. 是 $\hat{\mu}=6$.

367

(2) 一位试验者决定抽取观察值 $X_1,X_2,\cdots$ 直到她得到一个比10大的值. 她发现当 $i=1,\cdots,19$ 时都有 $X_i<10$，而 $X_{20}>10$. 因此，在观察到第20个观察值时抽样终止. 如果这20个观察值的平均数是6，那么 M. L. E. 又是 $\hat{\mu}=6$.

(3) 一位试验者头脑里并没有特别的计划，而且逐一取得观察值直到她被迫停止抽样或她感到她已有足够的信息来得到 μ 的一个好的估计. 假如由于任何终止原因她都收集了20个观察值，并且20个观察值的平均值是6，那么 M. L. E. 也是 $\hat{\mu}=6$.

有时，此种实验必须在试验者等待另一个顾客到来的区间内终止. 如果在最后一个顾客到达后又过去了很多时间，这段时间也不应该被删除在样本数据外. 假设例如，最先的20个观察值的平均数是6，实验者又等待了15分钟，但是没有其他顾客到来，那么她就结束了实验. 在此事例中，由于第21个观察值一定要比15大，尽管它的准确值未知，但我们可以知道 M. L. E. 应该大于6. 新的极大似然估计值可以通过把最先得到的20个观察值的似然函数与第21个观察值大于15的概率相乘来得到新的似然函数，并使新的似然函数达到极大来确定 θ 的值(见习题15).

记住似然函数决定极大似然估计值，M. L. E. 依赖于抽样方案唯一的途径就是通过似然函数. 如果关于什么时候停止观察数据完全由现在所观察到的观察值所决定，

那么该信息已经被包括在似然函数里了. 如果是由于其他原因而终止的，那么对于每个给定的 θ 值要将这“其他因素”的概率计算出来并要将此概率包括在似然中.

M. L. E. 的其他性质将会在第7章中再作讨论.

6.5.5 小结

参数 θ 的函数 g 的 M. L. E. 就是 θ 的 M. L. E. 的函数 g. 例如，如果 θ 是排队等候的顾客接受服务的比例，则 $1/\theta$ 是平均服务时间. $1/\theta$ 的 M. L. E. 就是1除以 θ 的 M. L. E. 有时我们找不到参数 M. L. E. 的一个接近的形式、表达式，当样本容量增大时，M. L. E. 序列依概率收敛于参数. 当在采集数据时何时终止仅取决于已经观察到的数据或基于与参数无关的其他一些考虑时，则该 M. L. E. 将不会依赖于抽样方案. 即如果两种不同的抽样方案导致成比例的似然函数，那么一个极大化似然函数的 θ 值也会使另一个极大化.

368

习题

1. 假设来自于概率密度函数如6.4节中习题9中所定义的一个分布的随机样本. 求 $e^{-1/\theta}$ 的 M. L. E. .
2. 假设从均值未知的泊松分布中随机抽取一个随机样本 $X_1\cdots,X_n$. 计算出 M. L. E. 的标准差.
3. 假设从参数 β 未知的指数分布中随机抽取一个随机样本 $X_1,\cdots X_n$. 计算 M. L. E. 的中位数.
4. 假设某种型号台灯的寿命服从参数 β 未知的指数分布. 在时间段 T 小时内，随机抽取 n 只该种型号的台灯做测试，并在此期间观察到失效的台灯数，但是没有记录下具体的失效时间. 计算基于 X 观察值的参数 β 的 M. L. E. .
5. 假设从区间为 $[a,b]$ 的均匀分布中随机抽取一个随机样本 $X_1,\cdots,X_n$，其中端点 a 和 b 都未知. 求出该分布均值的 M. L. E. .
6. 假设从均值和方差都未知的正态分布中随机抽取一个样本 $X_1,\cdots,X_n$. 找出此分布0.95分位点的 M. L. E.，即使得 $\Pr(X<\theta)=0.95$ 的 θ 的值.
7. 同样考虑习题6中的条件，找出 $v=\Pr(X>2)$ 的 M. L. E. .
8. 假设从密度函数如方程(6.5.2)中所定义的伽玛分布中随机抽取的一个样本 $X_1,\cdots,X_n$. 求 $\Gamma'(\alpha)/\Gamma(\alpha)$ 的 M. L. E. .
9. 假设从参数 α 和 β 都未知的伽玛分布中随机抽取一个样本 $X_1,\cdots,X_n$. 求 α/β 的 M. L. E. .
10. 假设从参数 α 和 β 均未知的 β 分布中随机抽取一个样本 $X_1\cdots X_n$. 求 α，β 的 M. L. E. 满足下列式子：

$$\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}-\frac{\Gamma'(\hat{\beta})}{\Gamma(\hat{\beta})}=\frac{1}{n}\sum_{i=1}^{n}\ln\frac{X_i}{1-X_i}.$$

11. 假设从区间为 $[0,\theta]$ 的均匀分布中抽取一个样本容量为 n 的随机样本 $X_1,\cdots,X_n$，其中 θ 的值未知. 求证：θ 的 M. L. E. 序列是一个相容序列.

12. 假设从参数β未知的指数分布中随机抽取一个样本$X_1,\cdots,X_n$. 证明：β的M.L.E.序列是一个相容序列.
13. 假设从概率密度函数如6.4节习题9中所定义的分布中随机抽取一个样本$X_1,\cdots,X_n$. 证明：θ的M.L.E.序列是一个相容序列.
14. 假设一位科学家希望估计女皇蝴蝶中有一种特殊花纹的比例p.
    a. 假设他逐个抓女皇蝴蝶，直到抓到5只有特殊花纹的蝴蝶. 如果他必须抓43只蝴蝶，那么p的M.L.E.是什么？
    b. 假设科学家在一天最终抓了58只女皇蝴蝶，而只发现了3只特殊花纹的蝴蝶，那么p的M.L.E.是什么？
15. 假设从均值μ未知的指数分布中随机抽取21个观察值($\mu>0$)，此20个观察值的平均数是6，虽然另外一个观察值的精确值未知，但可以确定它要大于15. 求μ的M.L.E..
16. 假设两位统计员A和B必须要估计某个参数θ且其值未知($\theta>0$)，A估计方法是观察一个服从参数为$\alpha=3$和$\beta=\theta$的伽玛分布的随机变量X的值；B估计方法是观察一个服从均值为2θ的泊松分布的随机变量Y的值. 如果A估计出来的值是$X=2$，而B估计的是$Y=3$. 证明：由这两个观察值所确定的似然函数成比例，并确定由每位统计员所得到的θ的M.L.E.的共同值. 369
17. 假设每个统计员A和B必须要估计某个未知参数$p(0<p<1)$，统计员A的估计方法是估计服从参数为$n=10$和p的二项分布的随机变量X的值；而统计员B是观察服从参数为$r=4$和p的负二项分布的随机变量Y的值. 假设统计员A得到的观察值是$X=4$，而统计员B得到的观察值是$Y=6$. 证明：由两个观察值所确定的似然函数是成比例的，并求由每个统计员所得到的p的M.L.E.的公共值. 370

第 7 章　估计量的抽样分布

7.1　统计量的抽样分布

统计量是一些可观察的随机变量的函数. 统计量的抽样分布告诉我们在得到观察数据之前，可以假设统计量取些什么值以及如何设定这些值. 称抽样分布为给定参数条件下统计量的分布.

7.1.1　统计量和估计量

在本书的许多例子中我们用随机样本的函数，像样本均值 $\overline{X}_n$ 和/或样本方差来估计参数 θ 或是参数 θ 的函数. 例如，在例 2.1.3 所介绍的诊疗试验中，我们可以用丙咪嗪组中观察到的没有旧病复发的病人的比例去估计参数“用丙咪嗪后没有旧病复发的病人的概率”. 在观察到数据之前，没有旧病复发的病人的比例是一个有分布的随机变量 X，而且它不可能恰好等于参数 θ. 然而，我们希望 X 接近于 θ 的可能性相当高. 比如，我们可以试着计算 $|X-\theta|<0.1$ 的概率. 这样的计算要求我们已经知道随机变量 X 的分布. 在诊疗试验中，我们可以设在丙咪嗪组中取 40 个病人，在给定 θ 条件下是条件独立的，其中 θ 是没有旧病复发的概率. 由此可知在给定 θ 的条件下 $40X$ 的条件分布是一个参数为 40 和 θ 的二项分布. 由此很容易得到 X 的分布. 事实上，X 的概率函数为：

391

$$f(x|\theta)=\binom{40}{40x}\theta^{40x}(1-\theta)^{40(1-x)},\text{对于 } x=0,\frac{1}{40},\frac{2}{40},\cdots,\frac{39}{40},1.$$

这种分布称为 X 的抽样分布，用它可以帮助我们回答在得到观察数据之前预期 X 接近于 θ 的程度之类的问题.

通常，假设的随机变量 $X_1,\cdots,X_n$ 是来自于一个含有未知参数 θ 的分布的随机样本. 定义统计量为随机变量 $X_1,\cdots,X_n$ 的任意一个实值函数 $T=r(X_1,\cdots,X_n)$. 由于统计量 T 是随机变量的函数，所以也是一个随机变量，且从理论上讲它的分布可以由 $X_1,\cdots,X_n$ 的联合分布得到，通常称该分布为是统计量 T 的抽样分布，因为它是由一个随机样本的观测值的联合分布得到的.

如 6.3 节中所提到的，估计量 $\delta(X_1,\cdots,X_n)$ 是随机变量 $X_1,\cdots,X_n$ 的一个函数，所以 θ 的估计量是一个统计量，因此，从理论上讲可以得到 θ 的每一个估计量的抽样分布. 实际上在本书的前几章中已经得到了很多估计量和统计量的分布. 例如，如果 $X_1,\cdots,X_n$ 是一个来自于均值是 μ、方差是 σ^2 的正态分布的随机样本，由 6.4 节可

知 μ 的极大似然估计是样本均值 $\overline{X}_n$，进而，在推论 5.6.2 可知 $\overline{X}_n$ 的分布是均值为 μ 方差为 σ^2/n 的正态分布.

在本章中对于来自于正态分布的随机样本，我们将得到样本方差的分布以及样本均值和样本方差的各种函数的分布. 这些结论将会使我们定义一些在统计推论问题中起部分重要作用的新分布. 另外，我们将研究某些估计量及其抽样分布的一般性质.

7.1.2 抽样分布的用途

从第 6 章讨论中可以看出，在一个给定问题中，可以在不用计算估计的抽样分布的情况下就确定贝叶斯估计或极大似然估计. 事实上，在得到观察值和找到参数 θ 的贝叶斯估计量后，该估计量的有关性质可以通过 θ 的后验分布来确定. 例如，估计量和未知量 θ 的差别在一个指定范围内的概率或是估计的最小二乘估计都可以由 θ 的后验分布来确定.

然而在抽样前，我们希望计算出统计量与未知量 θ 的差别不会超过某个范围的概率或计算估计量的最小二乘估计. 因此有必要对于 θ 的每个可能取值都要确定其估计量分布. 特别地，如果一个统计人员必需为得到 θ 的最佳估计量做以上两个或更多的试验中的某一个，或者是在一个指定的试验中选择最佳的样本容量，那么她可能使用不同估计量的抽样分布来做决定.

此外，如 6.2 和 6.3 节中所提到的，许多统计人员认为在某些问题中不可能或很难找到参数 θ 的一个较好的分布. 因此，在此类问题中也就不可能为参数 θ 确定一个后验分布. 所以，在观察到样本值和计算出 θ 估计量的数值后，就不可能或不再合适去考虑估计量接近于 θ 的后验概率. 在抽取样本前，统计人员可以利用估计量的抽样分布来计算估计量接近于 θ 的概率. 如果对于 θ 的每个可能取值该概率都是很高的，那么统计人员将推断她从样本观察值中得到的特定的估计量是非常接近于 θ，即便不能给出明确的后验分布. 然而，如我们将在例 7.5.4 的最后一个部分所说明的那样，不是总能保证能够得到这样的一个结论.

392

习题

1. 假设 $X_1,\cdots,X_n$ 是来自于区间 $[0,\theta]$ 均匀分布的一个随机样本，其中 θ 未知，取多大的样本才能使得对所有的 θ 都有 $\Pr\left(\left|\max\{X_1,\cdots,X_n\}-\theta\right|\leqslant 0.1\theta\right)\geqslant 0.95$?
2. 假设来自于一个均值 θ 未知，标准误差是 2 的正态分布的一个随机样本，取多大的样本才能使得对于每个可能的 θ 取值都有 $E_\theta\left(|\overline{X}_n-\theta|^2\right)\leqslant 0.1$?
3. 同习题 2 的条件，取多大的样本可以使得对于每个可能的 θ 都有 $E_\theta\left(\left|\overline{X}_n-\theta\right|\right)\leqslant 0.1$?
4. 同习题 2 的条件，取多大的样本可以使得对于每个可能的 θ，有 $\Pr\left(\left|\overline{X}_n-\theta\right|\leqslant 0.1\right)\geqslant 0.95$?
5. 假设一个随机样本来自于参数为 p 的伯努利分布，其中参数 p 的值未知. 并假设 p 的值落

在0.2的附近. 求取多大的样本才能使当 $p=0.2$ 时 $\Pr\left(\left|\bar{X}_n-p\right|\leqslant 0.1\right)\geqslant 0.75$?

6. 同习题5的条件，用第5.7节中的中心极限定理求出 $p=0.2$ 时 $\Pr\left(\left|\bar{X}_n-p\right|\leqslant 0.1\right)\geqslant 0.95$ 的近似随机样本容量?

7. 同习题5的条件，取多大的随机样本才能保证当 $p=0.2$ 时有 $\mathrm{E}_p\left(\left|\bar{X}_n-p\right|^2\right)\leqslant 0.01$?

8. 同习题5的条件，取多大的随机样本才能使得对于每个 p（$0\leqslant p\leqslant 1$）的每个可能值都有 $\mathrm{E}_p\left(\left|\bar{X}_n-p\right|^2\right)\leqslant 0.01$?

7.2　卡方分布

卡方(χ^2)分布族是Γ分布族一个子族. 该类特殊的Γ分布是针对从正态分布中抽取的随机样本的不同估计的抽样分布而产生的.

7.2.1　分布的定义

本节中我们将介绍和讨论Γ分布中的特殊一类，即卡方(χ^2)分布. 该分布与来自于正态分布中的随机样本紧密有关，在统计学领域中应用很广泛，在本书的其余部分我们可以看到它在统计推断的很多重要问题中的应用，在本节中我们将给出 χ^2 分布的定义以及它的一些基本数学性质.

393

在第5.9节中已经定义过参数为 α、β 的Γ分布. 对于每个正整数 n，称 $\alpha=n/2$ 且 $\beta=1/2$ 的Γ分布为自由度为 n 的 χ^2 分布. 若一个随机变量 X 服从自由度为 n 的 χ^2 分布，那么从(5.9.7)式可知，当 $x>0$ 时，X 的概率密度函数为

$$f(x)=\frac{1}{2^{n/2}\Gamma(n/2)}x^{(n/2)-1}\mathrm{e}^{-x/2}.\qquad(7.2.1)$$

并且，当 $x\leqslant 0$ 时，$f(x)=0$.

在本书后面给出了 χ^2 分布的对于 p 和 n 的不同取值的 p 分位数表. 大部分的统计软件包都包括具有计算任意 χ^2 分布的分布函数和分位数函数的功能.

从 χ^2 分布的定义及(7.2.1)式都能看出，自由度为2的 χ^2 分布是一个参数为1/2的指数分布，或等价于均值为2的一个指数分布. 因而以下三个分布是相同的：参数为 $\alpha=1$ 和 $\beta=1/2$ 的Γ分布，自由度为2的 χ^2 分布以及均值为2的指数分布.

7.2.2　分布的性质

如果随机变量 X 服从一个自由度为 n 的 χ^2 分布，如5.9节中所给出的Γ分布的均值和方差的表达式可以得到

$$\mathrm{E}(X)=n,\mathrm{Var}(X)=2n.\qquad(7.2.2)$$

进而，由5.9节中所给出的矩生成函数可以得到 X 的矩生成函数为

$$\psi(t)=\left(\frac{1}{1-2t}\right)^{n/2},\text{当 } t<\frac{1}{2}.$$

χ^2 分布的可加性，可以直接从定理5.9.3所得到，我们在下一个定理中不加证明地把它列出来.

定理7.2.1 若随机变量 $X_1,\cdots,X_k$ 是相互独立的，且每个 X_i 都服从自由度为 n_i 的 χ^2 分布 $(i=1,\cdots,k)$，那么和 $X_1+\cdots+X_k$ 服从自由度为 $n_1+\cdots+n_k$ 的 χ^2 分布. ■

现在我们来建立 χ^2 分布与正态分布的基本关系. 首先我们证明若一个随机变量 X 服从标准正态分布，则随机变量 $Y=X^2$ 将服从自由度为1的 χ^2 分布. 为了证明这一点，我们令 $f(y)$ 和 $F(y)$ 分别表示 Y 的概率密度函数和分布函数，同时，由于 X 服从标准正态分布. 我们令 $\phi(x)$ 和 $\Phi(x)$ 分别表示 X 的概率密度函数和分布函数. 则当 $y>0$ 时， [394]

$$\begin{aligned}F(y)&=\Pr(Y\leqslant y)=\Pr(X^2\leqslant y)=\Pr(-y^{1/2}\leqslant X\leqslant y^{1/2})\\&=\Phi(y^{1/2})-\Phi(-y^{1/2}).\end{aligned}$$

由于 $f(y)=F'(y)$ 且 $\phi(x)=\Phi'(x)$，由微分的链式法则可知

$$f(y)=\phi(y^{1/2})\left(\frac{1}{2}y^{-1/2}\right)+\phi(-y^{1/2})\left(\frac{1}{2}y^{-1/2}\right).$$

并且，因为 $\phi(y^{1/2})=\phi(-y^{1/2})=(2\pi)^{-1/2}e^{-y/2}$，可知

$$f(y)=\frac{1}{(2\pi)^{1/2}}y^{-1/2}e^{-y/2}\quad\text{当 } y>0 \text{ 时}.$$

将此等式与等式(7.2.1)作比较，可见 Y 的概率密度函数事实上就是自由度为1的 χ^2 分布的概率密度函数.

现在将该结果与定理7.2.1结合起来，得到以下的定理，此定理可以提供 χ^2 分布在统计中是一个重要分布的主要原因.

定理7.2.2 如果随机变量 $X_1,\cdots,X_k$ 独立同分布且均服从标准正态分布. 那么它们的平方和 $X_1^2+\cdots+X_k^2$ 服从一个自由度为 k 的 χ^2 分布. ■

例7.2.1 奶酪的乳酸浓度 Moore和McCabe(1999,p. D-1)描述了在澳大利亚所进行的一个为了研究口味和奶酪化学成分的关系的试验. 乳酸就是会影响口味的一种化学成分. 奶酪制造商如果想要拥有一批忠实的顾客，就会希望每个顾客在购买奶酪时味道尝上去会差不多. 化学成分例如乳酸的浓度就会导致奶酪口感的变化. 假定我们将一些大块奶酪中乳酸的浓度看作是均值为 μ 和方差为 σ^2 的独立正态分布变量，我们所感兴趣的是这些浓度和 μ 的值的相差程度. 令 $X_1,\cdots,X_k$ 是 k 个大块奶酪中的浓度，且令 $Z_i=(X_i-\mu)/\sigma$. 则

$$Y=\frac{1}{k}\sum_{i=1}^{k}|X_i-\mu|^2=\frac{\sigma^2}{k}\sum_{i=1}^{k}Z_i^2$$

可以度量 k 个浓度和 μ 相差的程度. 假设乳酸浓度的差别为 u 或者更高足以引起口味上有明显的差异. 那么我们希望去计算 $\Pr(Y\leqslant u^2)$. 根据定理7.2.2，$W=kY/\sigma^2$

的分布是自由度为 k 的 χ^2 分布．因此 $\Pr(Y\leqslant u^2)=\Pr(W\leqslant ku^2/\sigma^2)$．

例如，假设 $\sigma^2=0.09$，我们对10块奶酪感兴趣，此外，假定 $u=0.3$ 是我们感兴趣的临界差别点．我们可以写成

$$\Pr(Y\leqslant 0.3^2)=\Pr\left(W\leqslant\frac{10\times 0.09}{0.09}\right)=\Pr(W\leqslant 10). \tag{7.2.3}$$

使用自由度为10的 χ^2 分布的分位数表，可以看到10是界于0.5和0.6这两个分位数之间．事实上，可以用计算机软件来得到(7.2.3)式中的概率，其值是0.56，所以乳酸的浓度和这10大块奶酪的平均浓度的均方差就会有44%的可能性会大于预期的量．如果此概率过大，厂商就会希望在降低乳酸浓度的方差上花些功夫．◀

395

7.2.3　小结

自由度为 n 的卡方分布和参数为 $n/2$ 和 $1/2$ 的 Γ 分布是相同的．它是 n 个来自于标准正态分布且相互独立的随机变量的平方和的分布．自由度为 n 的 χ^2 分布的均值为 n，方差为 $2n$.

习题

1. 假设我们在例7.2.1中取20块奶酪作为样本，令 $T=\sum_{i=1}^{20}(X_i-\mu)^2/20$. 其中 X_i 表示第 i 块奶酪中乳酸的浓度，并设 $\sigma^2=0.09$，求 c 取何值会满足 $\Pr(T\leqslant c)=0.9$?
2. 确定自由度为 n 的 χ^2 分布的众数（$n=1, 2, \cdots$).
3. 画出自由度 n 分别取以下值时 χ^2 分布的概率密度函数，并在每个图上标明均值、中位数和众数．(a) $n=1$；(b) $n=2$；(c) $n=3$；(d) $n=4$.
4. 假定在 xy-平面中随机地选取一点 (X,Y)，其中 X 和 Y 是相互独立的随机变量，且都服从于标准正态分布．如果在 xy-平面内画一个圆心在原点的圆，求使点 (X,Y) 落在圆内的概率为0.99的圆半径最小是多少?
5. 假定在三维空间中随机选取一个点 (X,Y,Z)，其中 X，Y 和 Z 都相互独立且都服从标准正态分布．求从原点到该点的距离小于1的概率是多少?
6. 当我们在显微镜下观察液体或气体微粒的运动，看上去这个运动是不规则的，因为微粒之间时常相互碰撞．称该种运动的概率模型为布朗运动，定义如下：在液体或气体中选定一个坐标体系．假定在 $t=0$ 时微粒处在这个坐标体系中的原点，且令 (X,Y,Z) 表示在任意 $t>0$ 时刻时微粒的坐标，X，Y，Z 互相独立且都服从均值为0和方差为 σ^2t 的正态分布．确定在 $t=2$ 时微粒会落入以原点为球心，半径为 4σ 的球内的概率.
7. 假设随机变量 $X_1,\cdots,X_n$ 是相互独立的，而且每个随机变量 X_i 都有一个连续分布函数 F_i．令随机变量 Y 由关系式 $Y=-2\sum_{i=1}^{n}\ln F_i(X_i)$ 来确定．证明：Y 服从自由度为 $2n$ 的 χ^2 分布.
8. 假设 $X_1,\cdots,X_n$ 是来自于区间 $[0,1]$ 上均匀分布的一个随机样本，如3.9节中所定义的，令 W 表示样本的值域．令 $g_n(x)$ 表示随机变量 $2n(1-W)$ 的概率密度函数，且令 $g(x)$ 表

示自由度为4的χ^2分布的概率分布函数. 证明：

$$\lim_{n\to\infty} g_n(x) = g(x)，对于 x>0.$$

9. 假定$X_1,\cdots,X_n$是来自均值为μ和方差为σ^2的正态分布的一个随机样本. 确定$\frac{n(\overline{X}_n-\mu)^2}{\sigma^2}$的分布.
10. 假设6个来自于标准正态分布的随机变量$X_1,\cdots,X_6$组成一个随机样本，而令$Y=(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2$. 确定$c$的值使得随机变量$cY$服从$\chi^2$分布. [396]
11. 如果一个随机变量X服从自由度为n的χ^2分布，那么称$X^{1/2}$的分布为自由度为n的卡方(χ)分布，确定该分布的均值.
12. 再次考虑例7.2.1中的情形，求使得$\Pr(Y\leqslant 0.09)\geqslant 0.97$的最小的$\sigma^2$?

7.3 样本均值与样本方差的联合分布

假定我们的数据来自一个正态分布的随机样本. 为了估计正态分布的参数，样本均值$\hat{\mu}$和样本方差$\widehat{\sigma^2}$都是重要且必须计算出来的统计量，为了了解估计量的好坏，知道他们的联合分布也很有用处.

7.3.1 样本均值和样本方差的独立性

假定随机变量$X_1,\cdots,X_n$是来自于服从正态分布的随机样本，其中分布的均值μ未知且方差σ^2也未知. 那么，6.4节中所证明的μ和σ^2的极大似然估计分别是样本均值$\overline{X}_n$和样本方差$(1/n)\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$. 在本节中，我们要推导出这两个估计量的联合分布. [397]

从推论5.6.2可知，样本均值本身服从一个均值为μ，方差为σ^2/n的正态分布. 我们应证明一个值得注意的性质：样本均值和样本方差是互相独立的随机变量，虽然它们都是相同变量$X_1,\cdots,X_n$的函数. 而且我们还可以证明，除了一个比例系数，样本方差服从自由度为$n-1$的χ^2分布. 更准确地说，我们可以证明：随机变量$\sum_{i=1}^{n}(X_i-\overline{X}_n)^2/\sigma^2$服从自由度为$n-1$的$\chi^2$分布. 该结论也是来自于正态分布的随机样本的一个非常显著的性质，下面将深入讨论这一问题.

由于随机变量$X_1,\cdots,X_n$是互相独立的，又因为每个变量都服从均值为μ，方差为σ^2的正态分布，所以随机变量$(X_1-\mu)/\sigma,\cdots,(X_n-\mu)/\sigma$也是相互独立的，且每一个变量都服从标准正态分布. 由定理7.2.2可知，他们的平方和$\sum_{i=1}^{n}(X_i-\mu)^2/\sigma^2$服从自由度为$n$的$\chi^2$分布. 因此，正如前一段所提到的值得注意的该性质中，如果在平方和中用样本均值$\overline{X}_n$代替总体均值μ，它的影响仅仅是使χ^2分布的自由度由n变为$n-1$. 综上所述，我们可以得到以下定理：

定理7.3.1 假设$X_1,\cdots,X_n$是来自于服从均值为μ、方差为σ^2的正态分布的一个随

机样本，那么样本均值 $\overline{X}_n$ 和样本方差 $(1/n)\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 是相互独立的随机变量；$\overline{X}_n$ 服从均值为 μ，方差为 σ^2/n 的正态分布；$\sum_{i=1}^{n}(X_i-\overline{X}_n)^2/\sigma^2$ 服从自由度为 $n-1$ 的 χ^2 分布. ■

此外，还可以证明，仅当随机样本取自正态分布时，样本均值和样本方差才相互独立. 在本书中我们对这一结果不作深入研究. 然而，要强调的是：样本均值和样本方差的独立性的确是取自正态分布的样本的显著特性.

定理 7.3.1 的证明中用到了正交矩阵的性质，下面我们对此作介绍.

7.3.2　正交矩阵

1. 定义

称 $n\times n$ 阶矩阵 $\boldsymbol{A}$ 是正交的，若 $\boldsymbol{A}^{-1}=\boldsymbol{A}'$，其中 $\boldsymbol{A}'$ 是 $\boldsymbol{A}$ 的转置. 因此，矩阵 $\boldsymbol{A}'$ 是正交阵，当且仅当 $\boldsymbol{AA}'=\boldsymbol{A}'\boldsymbol{A}=\boldsymbol{I}$，其中 $\boldsymbol{I}$ 为 $n\times n$ 阶单位阵时. 由此定义得出：当且仅当每行元素的平方和为 1，并且不同两行中对应元素乘积的和是 0 时，矩阵才是正交的. 或者，当且仅当每列元素的平方和为 1，并且不同两列中对应元素乘积的和是 0 时，矩阵才是正交的.

2. 正交矩阵的性质

接下来我们将给出正交矩阵的两个重要性质.

性质 7.3.1　如果 $\boldsymbol{A}$ 是正交阵，那么 $|\det \boldsymbol{A}|=1$.

证明：为了证明结论，首先回顾一下，对于方阵 $\boldsymbol{A}$ 有 $\det\boldsymbol{A}=\det\boldsymbol{A}'$，且对于方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，有 $\det\boldsymbol{AB}=(\det\boldsymbol{A})(\det\boldsymbol{B})$. 因此，

$$\det(\boldsymbol{AA}')=(\det\boldsymbol{A})(\det\boldsymbol{A}')=(\det\boldsymbol{A})^2.$$

且，若 $\boldsymbol{A}$ 是正交阵，则 $\boldsymbol{AA}'=\boldsymbol{I}$. 所以

$$\det(\boldsymbol{AA}')=\det\boldsymbol{I}=1.$$

因此 $(\det\boldsymbol{A})^2=1$ 或等价地 $|\det\boldsymbol{A}|=1$. ■

性质 7.3.2　考虑二个 n 维随机向量

$$\boldsymbol{X}=\begin{bmatrix}X_1\\ \vdots\\ X_n\end{bmatrix}\quad 和\quad \boldsymbol{Y}=\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix},\tag{7.3.1}$$

假设 $\boldsymbol{Y}=\boldsymbol{AX}$，其中 $\boldsymbol{A}$ 是正交矩阵. 则

398

$$\sum_{i=1}^{n}Y_i^2=\sum_{i=1}^{n}X_i^2.\tag{7.3.2}$$

证明：此结论可以由等式 $\boldsymbol{AA}'=\boldsymbol{I}$ 直接得出. 因为

$$\sum_{i=1}^{n}Y_i^2=\boldsymbol{Y}'\boldsymbol{Y}=\boldsymbol{X}'\boldsymbol{A}'\boldsymbol{AX}=\boldsymbol{X}'\boldsymbol{X}=\sum_{i=1}^{n}X_i^2.$$ ■

向量 $\boldsymbol{X}$ 乘以一个正交矩阵 $\boldsymbol{A}$ 相当于 $\boldsymbol{X}$ 在 n 维空间中的旋转，可能会改变其中一

些坐标的符号. 以上任何这样的一些运算都不能改变原始向量 $\boldsymbol{X}$ 的长度，且此长度等于 $(\sum_{i=1}^{n} X_i^2)^{1/2}$.

综上所述，以上这两条正交矩阵的性质表明如果一个随机向量 $\boldsymbol{Y}$ 是通过随机向量 $\boldsymbol{X}$ 的正交线性变换 $\boldsymbol{Y}=\boldsymbol{AX}$ 得到的，那么该变换的雅可比行列式的绝对值为 1，而且 $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2$.

3. 随机样本的正交线性变换

假定 $X_1,\cdots,X_n$ 是来自于标准正态分布的一个随机样本. 那么 $X_1,\cdots,X_n$ 的联合概率密度函数如下，对于 $-\infty < x_i < \infty\ (i=1,\cdots,n)$ 有：

$$f_n(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right]. \tag{7.3.3}$$

假定矩阵 $\boldsymbol{A}$ 也是一个 $n\times n$ 阶的正交阵，随机变量 $Y_1,\cdots,Y_n$ 由关系式 $\boldsymbol{Y}=\boldsymbol{AX}$ 所确定，其中向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 由(7.3.1)式所确定. 这是一个线性变换，所以由(3.9.10)式可以得到 $Y_1,\cdots,Y_n$ 的联合概率密度函数

$$g_n(\boldsymbol{y}) = \frac{1}{|\det\boldsymbol{A}|}f_n(\boldsymbol{A}^{-1}\boldsymbol{y}).$$

令 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y}$. 因为 $\boldsymbol{A}$ 是正交阵，如我们刚才所证的 $|\det\boldsymbol{A}|=1$ 和 $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2$. 所以

$$g_n(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right). \tag{7.3.4}$$

从方程(7.3.4)我们可以看出 $Y_1,\cdots,Y_n$ 的联合概率密度函数和 $X_1,\cdots,X_n$ 的联合概率密度正好相同. 因此，我们可以得到以下结果：

定理 7.3.2 假如随机变量 $X_1,\cdots,X_n$ 是独立同分布且都服从于标准正态分布. 假定 $\boldsymbol{A}$ 是一个 $n\times n$ 阶的正交矩阵，以及 $\boldsymbol{Y}=\boldsymbol{AX}$. 则随机变量 $Y_1,\cdots,Y_n$ 也是独立同分布的，且也都服从于标准正态分布，而且 $\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2$. ■

7.3.3 证明定理 7.3.1：样本均值与样本方差的独立性

1. 来自标准正态分布的随机样本

首先我们将在 $X_1,\cdots,X_n$ 是一个来自于标准正态分布的随机样本的假设前提下来证明定理 7.3.1. 考虑 n 维行向量 $\boldsymbol{u}$，其中每一个分量都是 $1/\sqrt{n}$： [399]

$$\boldsymbol{u} = \left[\frac{1}{\sqrt{n}}\ \cdots\ \frac{1}{\sqrt{n}}\right]. \tag{7.3.5}$$

因为向量 $\boldsymbol{u}$ 的 n 个分量的平方和为 1，我们可能建立一个正交矩阵 $\boldsymbol{A}$ 使得向量 $\boldsymbol{u}$ 的分量作为矩阵 $\boldsymbol{A}$ 的第一行. 该构造，称为格莱姆-施密特正交化方法，在线性代数的教材中例如 Cullen(1972)中有描述，在这里不作讨论. 假定已构造这个矩阵 $\boldsymbol{A}$，我们

将再次以 $\boldsymbol{Y}=\boldsymbol{AX}$ 的变换来定义随机变量 $Y_1,\cdots,Y_n$.

因为以 $\boldsymbol{u}$ 的分量作为矩阵 $\boldsymbol{A}$ 的第一行，因而我们有

$$Y_1=\boldsymbol{uX}=\sum_{i=1}^{n}\frac{1}{\sqrt{n}}X_i=\sqrt{n}\,\overline{X}_n. \tag{7.3.6}$$

进而，通过定理 7.3.2，$\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}Y_i^2$. 因此得到

$$\sum_{i=2}^{n}Y_i^2=\sum_{i=1}^{n}Y_i^2-Y_1^2=\sum_{i=1}^{n}X_i^2-n\overline{X}_n^2=\sum_{i=1}^{n}(X_i-\overline{X}_n)^2.$$

从而我们可以获得下列等式

$$\sum_{i=2}^{n}Y_i^2=\sum_{i=1}^{n}(X_i-\overline{X}_n)^2. \tag{7.3.7}$$

从定理 7.3.2 可知随机变量 $Y_1,\cdots,Y_n$ 是相互独立的. 因此，随机变量 Y_1 与 $\sum_{i=2}^{n}Y_i^2$ 是独立的，接着可由等式(7.3.6)与(7.3.7)得出 $\overline{X}_n$ 和 $\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 是独立的. 而且，由定理 7.3.2 可知这 $n-1$ 个随机变量 $Y_2,\cdots,Y_n$ 是独立同分布的，且每个随机变量都是服从标准正态分布的. 因此，由定理 7.2.2 便知随机变量 $\sum_{i=2}^{n}Y_i^2$ 服从自由度为 $n-1$ 的 χ^2 分布. 由式(7.3.7)可推导出 $\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 也服从自由度为 $n-1$ 的 χ^2 分布.

2. 来自于任意一个正态分布的随机样本

在证明定理 7.3.1 的过程中，我们只考虑来自于标准正态分布的随机样本. 现在假设随机变量 $X_1,\cdots,X_n$ 是来自均值为 μ，方差为 σ^2 的一个任意正态分布的随机样本.

如果我们令 $Z_i=(X_i-\mu)/\sigma$，对于 $i=1,\cdots,n$，那么随机变量 $Z_1,\cdots,Z_n$ 是独立的，且每个变量都服从标准正态分布. 换言之，$Z_1,\cdots,Z_n$ 的联合分布与服从标准正态分布的随机变量的联合分布是一样的. 由刚得到的结果可知：$\overline{Z}_n$ 和 $\sum_{i=1}^{n}(Z_i-\overline{Z}_n)^2$ 是独立的，然而 $\sum_{i=1}^{n}(Z_i-\overline{Z}_n)^2$ 服从自由度为 $n-1$ 的 χ^2 分布. 由于

400

$$\sum_{i=1}^{n}(Z_i-\overline{Z}_n)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2. \tag{7.3.8}$$

因而可知样本均值 $\overline{X}_n$ 和样本方差 $(1/n)\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 是独立的，且等式(7.3.8)的右边的随机变量服从自由度为 $n-1$ 的 χ^2 分布. 至此命题 7.3.1 中所提出的结果都已证实.

例 7.3.1　人工降雨　Simpson，Olsen，和 Eden(1975)描述了这样一个实验：在该试验中我们考察用硝酸银人工产生的 26 朵云为一个随机样本所降的雨量是否比自然的 26 朵云为一个随机样本所降的雨量要多. 图 7-1 是来自于人为产生的云降雨量取对数后的直方图. 假设这些对数随机变量 $X_1,\cdots,X_{26}$ 是独立同分布的，且都服从均值为 μ、方差为 σ^2 的正态分布.

如果我们对于人工云降雨量的方差感兴趣的话，可以计算出样本方差$\widehat{\sigma^2}$ $\sum_{i=1}^{26}(X_i-\overline{X}_n)^2/26$. $U=26\,\widehat{\sigma^2}/\sigma^2$ 是服从自由度为 25 的χ^2 分布. 该分布可以告诉我们：对于不同的数量，$\widehat{\sigma^2}$高估或低估了 σ^2 的程度. 例如，本书中的χ^2 表告诉我们自由度为 25 的χ^2 分布的四分位数为 19.94，即 $\Pr(U\leqslant 19.94)=0.25$. 因此 [401]

$$0.25=\Pr\left(\frac{\widehat{\sigma^2}}{\sigma^2}\leqslant\frac{19.94}{26}\right)=\Pr(\widehat{\sigma^2}\leqslant 0.77\sigma^2). \qquad (7.3.9)$$

即，$\widehat{\sigma^2}$低估至少 23% 的 σ^2 的概率为 0.25. 在本例中$\widehat{\sigma^2}$的观察值为 2.460. 用等式(7.3.9)计算出的概率与观察值2.46 和 σ^2 的偏离程度是无关的. 等式(7.3.7)告诉我们的是$\widehat{\sigma^2}$低于 σ^2 至少 23% 的概率(在观察数据之前). ◀

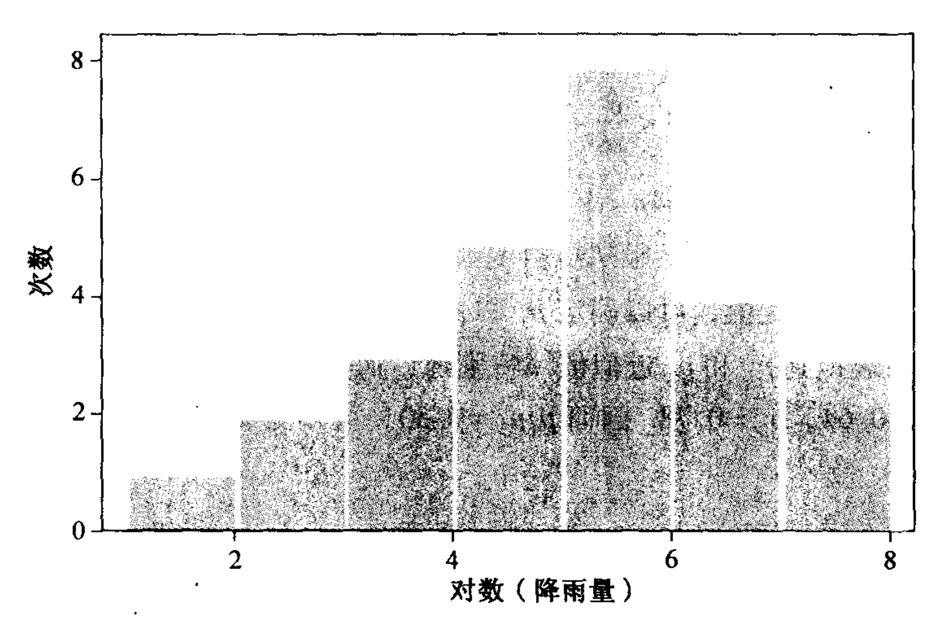


图 7-1 人工云对数降雨量的直方图

7.3.4 均值和标准差的估计

假定 $X_1,\cdots,X_n$ 是取自一个服从均值 μ 和标准差 σ 均未知的正态分布的随机样本. 同样，我们记 μ 和 σ 的极大似然估计为 $\hat{\mu}$ 和 $\hat{\sigma}$. 则

$$\hat{\mu}=\overline{X}_n \text{ 和 } \hat{\sigma}=\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right)^{1/2}.$$

注意：在 σ^2 的极大似然估计为 $\hat{\sigma}^2=\widehat{\sigma^2}$. 在本书的剩余部分中，一旦提到 σ^2 的极大似然估计时，我们可以用符号 $\hat{\sigma}^2$ 和$\widehat{\sigma^2}$中更简便的那个. 作为定理 7.3.1 应用的一个注释，我们要确定满足以下关系式的最小可能的样本容量 n：

$$\Pr\left(|\hat{\mu}-\mu|\leqslant\frac{1}{5}\sigma \text{ 和 } |\hat{\sigma}-\sigma|\leqslant\frac{1}{5}\sigma\right)\geqslant\frac{1}{2}. \qquad (7.3.10)$$

换言之，我们要确定 $\hat{\mu}$ 和 $\hat{\sigma}$ 与未知量的估计量差距都不会超过 $(1/5)\sigma$ 的概率至少为1/2的最小样本容量.

由于 $\hat{\mu}$ 和 $\hat{\sigma}$ 的独立性，关系式(7.3.10)可写为：

$$\Pr\left(|\hat{\mu}-\mu|\leqslant\frac{1}{5}\sigma\right)\Pr\left(|\hat{\sigma}-\sigma|\leqslant\frac{1}{5}\sigma\right)\geqslant\frac{1}{2}. \tag{7.3.11}$$

用 p_1 表示关系式(7.3.11)左边的第一个概率，令 U 为一个服从标准正态分布的随机变量，此概率可以表示成如下形式：

402

$$p_1=\Pr\left(\frac{\sqrt{n}\,|\hat{\mu}-\mu|}{\sigma}<\frac{1}{5}\sqrt{n}\right)=\Pr\left(|U|<\frac{1}{5}\sqrt{n}\right).$$

类似地，记 p_2 为关系式(7.3.11)左边第二个概率，且设 $V=n\hat{\sigma}^2/\sigma^2$，该概率可写为以下形式：

$$\begin{aligned}p_2&=\Pr\left(0.8<\frac{\hat{\sigma}}{\sigma}<1.2\right)=\Pr\left(0.64n<\frac{n\hat{\sigma}^2}{\sigma^2}<1.44n\right)\\&=\Pr(0.64n<V<1.44n).\end{aligned}$$

由定理7.3.1可得随机变量 V 服从自由度为 $n-1$ 的 χ^2 分布.

对于 n 的每个特定值，可以确定 p_1 和 p_2 的值，至少也可以通过本书最后标准正态分布表和 χ^2 表得到 p_1 和 p_2 近似值. 特别地，通过试过各种各样的 n 值后，发现当 $n=21$ 时，$p_1=0.64$，$p_2=0.78$. 因而 $p_1p_2=0.50$，可知当 $n=21$ 时，关系式(7.3.10)是成立的.

7.3.5　小结

令 $X_1,\cdots,X_n$ 是来自于服从均值为 μ，方差为 σ^2 的正态分布的一个随机样本. 则该样本的均值 $\hat{\mu}=\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$，样本方差 $\widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 是相互独立随机变量. 进而 $\hat{\mu}$ 服从均值为 μ，方差为 σ^2/n 的正态分布，且 $n\widehat{\sigma^2}/\sigma^2$ 服从自由度为 $n-1$ 的卡方分布.

习题

1. 假设 $X_1,\cdots,X_n$ 是来自于均值为 μ，方差为 σ^2 的正态分布的一个随机样本. 证明：$\widehat{\sigma^2}$ 服从参数为 $(n-1)/2$ 和 $n/(2\sigma^2)$ 的 Γ 分布.
2. 判断以下5个矩阵中哪几个是正交阵.

a. $\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}$　　b. $\begin{bmatrix}0.8&0&0.6\\-0.6&0&0.8\\0&-1&0\end{bmatrix}$　　c. $\begin{bmatrix}0.8&0&0.6\\-0.6&0&0.8\\0&0.5&0\end{bmatrix}$

d. $\begin{bmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$,　　e. $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$

3. a. 构造一个 2×2 阶的正交矩阵，其第一行的数据为：

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

b. 构造一个 3×3 阶的正交矩阵，其第一行为：

$$\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

403

4. 假设随机变量 X_1, X_2, X_3 独立同分布，且每个都服从标准正态分布．并且假设

$$Y_1 = 0.8X_1 + 0.6X_2,$$
$$Y_2 = \sqrt{2}(0.3X_1 - 0.4X_2 - 0.5X_3),$$
$$Y_3 = \sqrt{2}(0.3X_1 - 0.4X_2 + 0.5X_3).$$

请找出 Y_1, Y_2, Y_3 的联合分布.

5. 假设随机变量 X_1 和 X_2 是独立的，且每个都服从均值为 μ，方差为 σ^2 的正态分布．证明：随机变量 X_1+X_2 和 X_1-X_2 也是独立的.

6. 假设 $X_1,\cdots,X_n$ 是来自于服从均值为 μ，方差为 σ^2 的正态分布的一个随机样本．假设该样本容量 n 为 16，确定下列概率的值：

a. $\Pr\left[\frac{1}{2}\sigma^2 \leqslant \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \leqslant 2\sigma^2\right]$, b. $\Pr\left[\frac{1}{2}\sigma^2 \leqslant \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2 \leqslant 2\sigma^2\right]$.

7. 假设 $X_1,\cdots,X_n$ 是来自于服从均值为 μ、方差为 σ^2 的正态分布的一个随机样本，并令 $\widehat{\sigma^2}$ 为样本方差．求满足下列关系的 n 的最小值：

a. $\Pr\left(\frac{\widehat{\sigma^2}}{\sigma^2} \leqslant 1.5\right) \geqslant 0.95$;　　b. $\Pr\left(|\widehat{\sigma^2}-\sigma^2| \leqslant \frac{1}{2}\sigma^2\right) \geqslant 0.8$.

8. 假设 X 是服从自由度为 200 的 χ^2 分布. 请解释为什么可以用中心极限定理来确定 $\Pr(160 < X < 240)$ 的近似值，并确定此近似值.

9. 假设两个统计人员 A 和 B 独立地从服从均值 μ 未知，方差 $\sigma^2=4$ 的正态分布中选取 20 个观察值作为随机样本. 假设统计人员 A 发现他的样本方差是 3.8. 而统计人员 B 发现他的样本方差是 9.4. 那么谁的样本均值更接近于 μ 的未知值.

7.4 t 分布

当我们的数据是来自于服从均值为μ、方差为σ^2的正态分布的一个样本，$Z = n^{1/2}(\hat{\mu}-\mu)/\sigma$的分布是标准正态分布，其中$\hat{\mu}$是样本均值. 如果$\sigma^2$是未知的，我们可以用一个估计值（特别地，$\sigma' = [n\widehat{\sigma^2}/(n-1)]^{1/2}$）来代替$Z$的公式中的$\sigma$，所得的该随机变量服从自由度为$n-1$的$t$分布，且对于单独推断$\mu$是十分有用的，即使$\mu$和$\sigma^2$都未知的情况下也成立.

7.4.1 分布的定义

在本节中我们将介绍和讨论另一种分布，叫作t分布，它与服从正态分布的随机样本有着十分密切的关系. t分布和χ^2分布一样，在统计推断重要问题中有广泛应用. t分布也叫作学生分布(见 Student, 1908)为了记念 W. S. Gosset，他在 1908 年用笔名“Student”发表了对该分布的研究. 该分布定义如下：

考虑两个独立随机变量Y和Z，其中Y服从自由度为n的χ^2分布，而Z服从标准正态分布. 假设一个随机变量X定义为

404

$$X = \frac{Z}{\left(\frac{Y}{n}\right)^{1/2}}. \tag{7.4.1}$$

那么就称X的分布是自由度为n的t分布.

1. 概率密度函数的推导

我们来推导出自由度为n的t分布的概率密度函数. 假设Y与Z的联合分布同t分布中的定义形式相同. 那么由于Y与Z是相互独立的，因而它们的联合概率密度函数等于$f_1(y)$与$f_2(z)$的乘积，其中$f_1(y)$是自由度为n的χ^2分布的概率密度函数，而$f_2(z)$是标准正态分布的概率密度函数. 令X如等式(7.4.1)中所定义，为了简便起见，令$W=Y$. 我们可以首先确定X和W的联合概率密度函数.

由X与W的定义可知，

$$Z = X\left(\frac{W}{n}\right)^{1/2} \text{并且 } Y = W. \tag{7.4.2}$$

从X和W到Y和Z的变换(7.4.2)的雅可比变换是$(W/n)^{1/2}$. X和W的联合概率密度函数$f(x,w)$可以通过使用(7.4.2)中给出的式子来替代联合概率密度函数$f_1(y)f_2(z)$中的y和z，然后将结果乘以$(w/n)^{1/2}$来获得. 然后可以发现$f(x,w)$的值，具体步骤如下：当$-\infty < x < \infty$且$w>0$时：

$$f(x,w) = f_1(w)f_2\left(x\left[\frac{w}{n}\right]^{1/2}\right)\left(\frac{w}{n}\right)^{1/2} = cw^{(n+2)/2-1}\exp\left[-\frac{1}{2}\left(1+\frac{x^2}{n}\right)w\right], \tag{7.4.3}$$

其中 $c=\left[2^{(n+1)/2}(n\pi)^{1/2}\Gamma\left(\frac{n}{2}\right)\right]^{-1}$.

X 的边际概率密度函数 $g(x)$ 可由等式(7.4.3)通过以下关系获得

$$g(x)=\int f(x,w)\mathrm{d}w=c\int_0^{\infty}w^{(n+1)/2-1}\exp[-wh(x)]\mathrm{d}w,$$

这里 $h(x)=[1+x^2/n]/2$. 由等式(5.9.8)得:

$$g(x)=c\frac{\Gamma((n+1)/2)}{h(x)^{(n+1)/2}}=\frac{\Gamma(\frac{n+1}{2})}{(n\pi)^{1/2}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2},-\infty<x<\infty. \quad (7.4.4)$$

405

因此，如果 X 服从自由度为 n 的 t 分布，那么它的概率密度函数就由等式(7.4.4)所确定.

2. 与柯西分布及正态分布的关系

从等式(7.4.4)中可以看出概率密度函数 $g(x)$ 是对称的钟形函数，它的最大值在 $x=0$ 时取到. 因此，它一般的图形与均值为 0 的正态分布的概率密度函数很相似. 然而，当 $x\to\infty$ 或 $x\to-\infty$ 时，概率密度函数 $g(x)$ 尾端比一个正态分布概率密度函数尾端趋于零的速度要慢得多. 事实上，从等式(7.4.4)可以看出，当 $n=1$ 时，t 分布即成为柯西分布，该内容在 4.1 节中已描述过. 柯西分布的概率密度函数在图 4-3 中已有大致的图像. 在 4.1 节中已证明柯西分布的均值不存在，因为用来确定均值的积分不是绝对收敛的. 因而可知尽管自由度为 1 的 t 分布的概率密度函数关于 $x=0$ 点对称，但是这个分布的均值不存在.

同样由等式(7.4.4)可以证明，当 $n\to\infty$ 时，概率密度函数 $g(x)$ 对于 $-\infty<x<\infty$ 中的每个值 x 都收敛到标准正态分布的概率密度函数 $\phi(x)$. 该结论可以从以下两个极限结果得到:

$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^{bn}=\mathrm{e}^{ab}, \quad (7.4.5)$$

$$\lim_{m\to\infty}\frac{\Gamma(m+\frac{1}{2})}{\Gamma(m)m^{1/2}}=1.$$

(参见习题 7,找一个方法证明第一个等式.)由此，当 n 很大时，自由度为 n 的 t 分布可用标准正态分布来近似. 图 7-2 显示了标准正态分布的概率密度与自由度分别为 1、5 和 20 的 t 分布的概率密度函数，这样读者就可以看到当自由度增加时，t 分布是如何靠近正态分布的.

406

本书末尾给出了 t 分布关于 p 和 n 不同取值时 p 分位数的表. 表格中首行的概率与 $n=1$ 相对应，即与柯西分布相对应. 表的末行概率与 $n=\infty$ 相对应，即与标准正态分布相对应. 多数统计包都具有计算任意一个 t 分布的分布函数和分位数函数的功能.

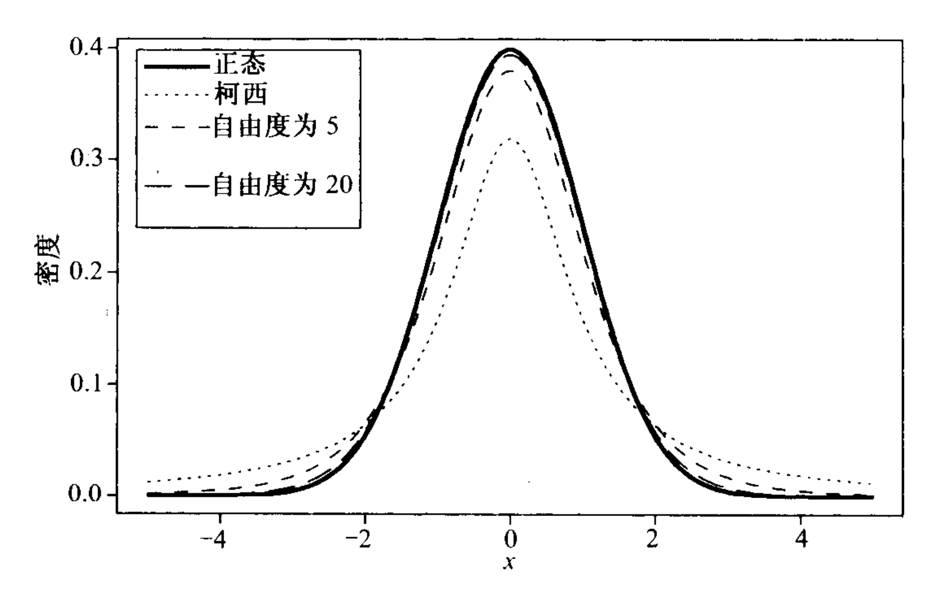


图 7-2　标准正态分布和 t 分布的概率密度函数

7.4.2　t 分布的矩

虽然当 $n=1$ 时 t 分布的均值不存在，但是当 n 取大于 1 的任意一个值时，均值都是存在的. 当然，只要均值存在，它的值就是 0. 这是因为 t 分布的对称性.

一般地，如果一个随机变量 X 服从自由度为 $n(n>1)$ 的 t 分布，那么就有 $E(|X|^k)<\infty$，当 $k<n$ 时；$E(|X|^k)=\infty$，当 $k\geqslant n$ 时. 即，X 的前 $n-1$ 阶矩存在，但是更高阶的矩不存在. 由此可知 X 的矩母函数也不存在.

可以证明（参见本节最后的习题 1），如果 X 服从一个自由度为 $n(n>2)$ 的 t 分布，那么 $\mathrm{Var}(X)=n/(n-2)$.

7.4.3　与来自正态分布的随机样本的关系

再次假设 $X_1,\cdots,X_n$ 是一组服从均值为 μ、方差为 σ^2 的正态分布的随机样本. 和以往一样，我们用 $\overline{X}_n$ 来表示样本均值，并且令

$$S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

如果我们定义随机变量 Y 和 Z，且有 $Z=n^{1/2}(\overline{X}_n-\mu)/\sigma$ 及 $Y=S_n^2/\sigma^2$，那么根据定理 7.3.1 可得 Y 和 Z 是独立的，Y 服从自由度为 $n-1$ 的 χ^2 分布，Z 服从标准正态分布. 现在考虑另一个随机变量 U，将它定义成

$$U=\frac{Z}{\left(\dfrac{Y}{n-1}\right)^{1/2}}.$$

根据 t 分布的定义，U 服从自由度为 $n-1$ 的 t 分布，可以很容易地得出 U 的表达式又可以写成如下形式：

$$U=\frac{n^{1/2}(\overline{X}_n-\mu)}{\left(\frac{S_n^2}{n-1}\right)^{1/2}}. \tag{7.4.6}$$

407

是 R. A. 费希尔在 1923 年第一次严格地证明了 U 分布是自由度为 $n-1$ 的 t 分布.

在(7.4.6)式中很重要的一点就是 U 的值和 U 分布都和方差 σ^2 的取值无关. 我们之前已经提到随机变量 $Z=n^{1/2}(\overline{X}_n-\mu)/\sigma$ 是服从标准正态分布的. 如果该表达式中标准差 σ 用一个它的估计值来代替，该估计值是通过样本 $X_1,\cdots,X_n$ 计算得出的，那么 Z 的分布就发生改变了. 特别地，如果我们用 σ 的极大似然估计 $\hat{\sigma}=(S_n^2/n)^{1/2}$ 来代替它，那么我们得到随机变量 Z'，它的定义如下：

$$Z'=\frac{n^{1/2}(\overline{X}_n-\mu)}{\hat{\sigma}}=\left(\frac{n}{n-1}\right)^{1/2}U.$$

由于 Z' 和 U 仅仅相差一个常量，Z' 的分布可以很容易地从 U 的分布中衍生而得，而已经得知 U 的分布是一个自由度为 $n-1$ 的 t 分布. 所以，如果标准差 σ 用它的极大似然估计 $\hat{\sigma}$ 来代替，则所得的随机变量 Z' 以及它的分布就不再依赖于正态分布下的方差 σ^2. 该事实在实际应用中的重要性会在本节的后面以及本书的其他一些章节中给出说明.

关于随机变量 U 本身有一个类似的陈述. 在 Z 的表达式中，我们用估计量 σ' 来代替标准差 σ.

$$\sigma'=\left[\frac{S_n^2}{n-1}\right]^{1/2}=\left(\frac{n}{n-1}\right)^{1/2}\hat{\sigma}. \tag{7.4.7}$$

那么随机变量 Z 将会被随机变量 U 代替. 我们可以从(7.4.7)式中发现当 n 取较大值时，估计量 σ' 和 σ 两者将会非常接近，将在第 7.7 节中进一步讨论估计量 σ'.

如果样本容量 n 很大，那么估计量 σ' 和 σ 很接近的概率就很高. 因此，在随机变量 Y 中用 σ' 代替 σ，不会很大地改变 Z 的标准正态分布. 由于该原因，就像我们之前提到过的，当 $n\to\infty$ 时，t 分布的概率分布函数收敛于标准正态分布.

例 7.4.1 人工降雨 采用和例 7.3.1 中相同的样本，考虑当 $n=26$ 时对数降雨量的测量. 假设现在我们想要知道这些测量值的样本平均值 $\overline{X}_n$ 和均值 μ 相差多远. 假定观察值 $X_1,\cdots,X_n$(对数降雨量)是相互独立的，且服从同一正态分布. $U=n^{1/2}(\overline{X}_n-\mu)/\sigma'$ 的分布就是一个自由度为 $n-1$ 的 t 分布. 对于 $n=26$，t 分布表告诉我们当自由度为 25 时，0.9 分位数的值是 1.316. 所以 $\Pr(U\leqslant 1.316)=0.9$. 也就有下式

$$\Pr(\overline{X}_n\leqslant\mu+0.2581\sigma')=0.9,$$

因为 $1.316/(26)^{1/2}=0.2581$. 即，$\overline{X}_n$ 不会超过 μ 的 $0.2581\sigma'$ 倍的概率为 0.9. 当然，σ' 和 $\overline{X}_n$ 都是随机变量，所以该结果不如我们所希望的那么可靠. 在 7.5 和 7.6 节中我们会说怎样利用 t 分布对于未知的均值 μ 做一些标准推断. ◀

408

7.4.4 小结

设 $X_1,\cdots,X_n$ 是来自于均值为 μ、方差为 σ^2 的正态分布的一个随机样本. 设 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i, \sigma'=\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right)^{1/2}$. 则 $n^{1/2}(\overline{X}_n-\mu)/\sigma'$ 的分布是自由度为 $n-1$ 的 t 分布.

习题

1. 假设 X 服从自由度为 $n(n>2)$ 的 t 分布. 证明: $\mathrm{Var}(X)=n/(n-2)$.
   提示: 为了计算 $E(X^2)$, 只需在正半实数轴上求积分且把变量 x 变换成
   $$y=\frac{\frac{x^2}{n}}{1+\frac{x^2}{n}}.$$
   然后将所得积分与 β 分布的概率密度函数作比较. 另一种方法, 利用 5.9 节中的习题 21.
2. 设 $X_1,\cdots,X_n$ 是来自于均值 μ 和标准差 σ 都未知的正态分布的随机样本, 分别令 $\hat{\mu}$ 和 $\hat{\sigma}$ 表示 μ 和 σ 的极大似然估计, 假设样本容量 $n=17$, 找到一个 k 值, 使得
   $$\Pr(\hat{\mu}>\mu+k\hat{\sigma})=0.95.$$
3. 假设五个随机变量 $X_1,\cdots,X_5$ 独立同分布, 且每个均服从标准正态分布. 确定一个常数 c, 使随机变量
   $$\frac{c(X_1+X_2)}{(X_3^2+X_4^2+X_5^2)^{1/2}}$$
   服从 t 分布.
4. 用书后给出的 t 分布表, 计算积分
   $$\int_{-\infty}^{2.5}\frac{\mathrm{d}x}{(12+x^2)^2}$$
   的值.
5. 假设随机变量 X_1 和 X_2 互相独立, 并均服从均值为 0, 方差为 σ^2 的正态分布. 计算
   $$\Pr\left[\frac{(X_1+X_2)^2}{(X_1-X_2)^2}<4\right].$$
   提示: $(X_1-X_2)^2=2\left[\left(X_1-\frac{X_1+X_2}{2}\right)^2+\left(X_2-\frac{X_1+X_2}{2}\right)^2\right].$
6. 在例题 7.2.1 中, 假设我们观察到 $n=20$ 块奶酪的乳酸含量为 $X_1,\cdots,X_{20}$. 找到一个数 c, 使得 $\Pr(\overline{X}_{20}\leqslant\mu+c\sigma')=0.95$.
7. 证明等式(7.4.5)中的极限公式. 提示: 等式两边分别取对数, 对于较小的 x, 用 $\ln(1+x)$ 的展开式来表示.

8. 设X服从标准正态分布，Y服从自由度为5的t分布. 说明为什么$c=1.63$时，$\Pr(-c<X<c)-\Pr(-c<Y<c)$的差值能取到最大值. 提示：从图7-2开始入手.

7.5 置信区间

置信区间为我们提供了当希望估计未知参数θ时使用估计值$\hat{\theta}$的另一种方法. 我们找到一个区间(A,B)，使其能以较大的概率包含θ. 区间长度可以让我们了解估计θ的近似程度.

409

7.5.1 正态分布均值的置信区间

我们仍假设$X_1,\cdots,X_n$为来均值μ未知、方差σ^2也未知的正态分布的一个随机样本. 令$t_{n-1}(x)$代表自由度为$n-1$的t分布的概率密度函数，设c是使

$$\int_{-c}^{c} t_{n-1}(x)\,\mathrm{d}x = \gamma. \tag{7.5.1}$$

成立的一个常数. 对于n的每个值，都能在本书最后的t分布表中找到相应的c值. 例如：若$n=12$，$T_{11}(x)$代表自由度为11的t分布的分布函数，则

$$\int_{-c}^{c} t_{11}(x)\,\mathrm{d}x = T_{11}(c) - T_{11}(-c) = T_{11}(c) - [1 - T_{11}(c)] = 2T_{11}(c) - 1.$$

如果$\gamma=0.95$，由等式(7.5.1)可得$T_{11}(c)=0.975$. 经查表得$c=2.201$为自由度为11的t分布的0.975百分位点.

根据等式(7.4.6)中所定义的随机变量U服从自由度为$n-1$的t分布，所以等式(7.5.1)意味着$\Pr(-c<U<c)=\gamma$. 进而，由等式(7.4.6)和(7.4.7)可以将该关系式重新写为：

$$\Pr\left(\overline{X}_n - \frac{c\sigma'}{n^{1/2}} < \mu < \overline{X}_n + \frac{c\sigma'}{n^{1/2}}\right) = \gamma. \tag{7.5.2}$$

因此，等式(7.5.2)表明虽然σ的值未知，但μ介于随机变量$A=\overline{X}_n-(c\sigma'/n^{1/2})$，$B=\overline{X}_n+(c\sigma'/n^{1/2})$之间的概率为$\gamma$.

在实际问题中，等式(7.5.2)应用如下；在随机样本中，若观察到变量$X_1,\cdots,X_n$的值，则能通过计算得到A和B的值. 如果$A=a$且$B=b$，则称区间(a,b)为μ的置信度为γ或$100\gamma\%$的置信区间. 因此我们就很自然地认为，未知量μ以γ的置信度落在区间(a,b)中.

例7.5.1 人工降雨 在例题7.3.1中，人工云的对数降雨量的均值$\overline{X}_n=5.134$. σ'的观察值为1.600. 设$\gamma=0.95$. 书后的t分布给出了当自由度为25时，0.975分位点的值为$c=2.060$. 所以置信度为95%的置信区间的上下限分别为$5.134-2.060\times 1.600/(26)^{1/2}=4.488$和$5.134+2.060\times 1.600/(26)^{1/2}=5.780$. ◀

例7.5.2 奶酪中乳酸的含量 在例题7.2.1中，我们讨论了样本容量为10的奶酪中乳酸的含量. 假设我们要计算出未知量μ的90%置信区间，其中μ为未知的乳酸的平均含量. 在等式(7.5.1)中当$n=10$，$\gamma=0.9$时，即t分布自由度为9时0.95分位点$c=1.833$. 由等式(7.5.2)可知，置信区间的端点为$\overline{X}_n \pm 1.833\sigma'/(10)^{1/2}$. 以下是由Moore和McCabe(1999, p. D-1)得到的10块奶酪中乳酸含量的观察值，数据如下：

410

0.86，1.53，1.57，1.81，0.99，1.09，1.29，1.78，1.29，1.58

这10个数据的平均值$\bar{x}_n=1.379$, $\sigma'=0.3277$. 所以，得到90%的置信区间的端点为$1.379-1.833\times0.3277/(10)^{1/2}=1.189$和$1.379+1.833\times0.3277/(10)^{1/2}=1.569$.

对于一个给定的置信度，对于μ可以构造出很多不同的置信区间. 例如，假设置信度为0.95，对于每对使得$\Pr(c_1<U<c_2)=0.95$的常数c_1和c_2，可以构造出μ的两个端点分别为

$$a=\bar{x}_n-\frac{c_2\sigma'}{n^{1/2}}, b=\bar{x}_n-\frac{c_1\sigma'}{n^{1/2}}$$

的置信区间.

然而，可以证明，在所有置信度为0.95的置信区间中，由等式(7.5.2)得出关于$\bar{x}_n$值对称的区间是区间长度最短的.

7.5.2 任意参数的置信区间

一般情况下，假设$X_1,\cdots,X_n$是一个来自于包含未知参数θ的分布的随机样本. 同时假设无论θ的真值可能是什么，都可以被找到两个统计量$A(X_1,\cdots,X_n)$和$B(X_1,\cdots,X_n)$，使得

$$\Pr[A(X_1,\cdots,X_n)<\theta<B(X_1,\cdots,X_n)]=\gamma, \tag{7.5.3}$$

其中γ是一个固定的概率（$0<\gamma<1$）. 如果$A(X_1,\cdots,X_n)$和$B(X_1,\cdots,X_n)$的观察值是a和b，那么就说区间(a,b)是θ的以γ为置信度的一个置信区间，即，θ是以置信度γ落在区间(a,b)中.

注：置信区间的另一种定义. 许多作者正如我们前面那样来定义置信区间. 但也有许多作者将置信区间定义为随机区间$(A(X_1,\cdots,X_n),B(X_1,\cdots,X_n))$，他们称$(a,b)$为置信区间的观察值，用此方法的目的是为了表明随机变量与它的观察值是有区别的. 贯穿本书，我们将继续使用我们已经给出的定义，但是要进一步研究统计的读者可能在以后会遇到另外一种定义. 有些作者还会用闭区间取代开区间来定义置信区间.

例7.5.3 泊松分布参数的近似置信区间 假设$X_1,\cdots,X_n$来自于一个含有未知参数θ的泊松分布. 假设n足够大，以致于$\overline{X}_n$近似地有一个正态分布. 在例题5.7.5中，我们发现

$$\Pr\left(|2\overline{X}_n^{1/2}-2\theta^{1/2}|<c\right)\approx 2\Phi(cn^{1/2})-1. \tag{7.5.4}$$

411

在我们观察到 $\overline{X}_n=x$，(7.5.4)式告诉我们

$$(-c+2x^{1/2},c+2x^{1/2}) \tag{7.5.5}$$

是 $2\theta^{1/2}$ 的置信度为 $2\Phi(cn^{1/2})-1$ 的一个近似置信区间. 例如，如果 $c=0.196$ 及 $n=100$，那么 $2\Phi(cn^{1/2})-1=0.95$. $g(\theta)=2\theta^{1/2}$ 的反函数 $g^{-1}(y)=y^2/4$，当 $y\geq 0$，这是一个递增的函数. 如果(7.5.5)式中的两个端点都是非负的，那么我们可知 $2\theta^{1/2}$ 落在(7.5.5)式的区间里当且仅当 θ 在区间

$$\left(\frac{1}{4}[-c+2x^{1/2}]^2,\frac{1}{4}[c+2x^{1/2}]^2\right). \tag{7.5.6}$$

如果 $-c+x^{1/2}<0$，(7.5.5)和(7.5.6)式中的左端点应该被 0 所取代. 经过该修正，(7.5.6)是 θ 的一个置信度为 $2\Phi(cn^{1/2})-1$ 的近似置信区间. ◀

7.5.3 扩展研究：置信区间的缺点

1. 置信区间的说明

应该要强调说明的是 θ 以概率 γ 落入区间 (a,b) 的说法是不正确的. 我们要在这里对这点作更进一步的说明. 在观察到统计量 $A(X_1,\cdots,X_n)$ 和 $B(X_1,\cdots,X_n)$ 的值之前，这些统计量都是随机变量. 因此，根据等式(7.5.3)下，θ 以概率 γ 落入以 $A(X_1,\cdots,X_n)$ 和 $B(X_1,\cdots,X_n)$ 为端点的随机区间中，但在观察到 $A(X_1,\cdots,X_n)=a$，$B(X_1,\cdots,X_n)=b$ 这一特定值之后，来考虑 θ 落入该特定区间 (a,b) 内事件的概率，且不把 θ 看作是一个随机变量，但该变量本身却有一个概率分布，这是不可能的. 即，必须给 θ 指定一个先验分布，然后用所得的后验分布去计算 θ 在区间 (a,b) 内的概率. 许多统计学家不给参数 θ 指定一个先验分布，更倾向于说 θ 以置信度 γ 而不是以概率 γ 落入区间 (a,b). 正是因为置信度和概率之间的这种差别，置信区间的含义及其相关的内容都是统计实践中有些争议的话题.

2. 可被忽略的信息

为了与先前的解释一致，置信区间的置信度 γ 的说明如下：在得到一个样本之前，由样本所构造的区间包括未知值 θ 的概率为 γ. 在样本值被观察到之后，可能就会有额外信息来帮助我们判断从这些数据所得到的区间是否真的包含 θ. 而如何利用该信息来调整置信度 γ 则是另一个有争议的话题.

例 7.5.4 长度为 1 的区间的相似 假设从区间 $\left[\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right]$ 上的一均匀分布中随机抽取两个观察值 X_1 和 X_2，其中 θ 未知，且 $-\infty<\theta<\infty$. 如果我们让 $Y_1=\min(X_1,X_2)$，$Y_2=\max(X_1,X_2)$，于是有

412

$$
\begin{aligned}
\Pr(Y_1 < \theta < Y_2) &= \Pr(X_1 < \theta < X_2) + \Pr(X_2 < \theta < X_1) \\
&= \Pr(X_1 < \theta)\Pr(X_2 > \theta) + \Pr(X_2 < \theta)\Pr(X_1 > \theta) \\
&= (1/2)(1/2) + (1/2)(1/2) = 1/2.
\end{aligned} \tag{7.5.7}
$$

根据(7.5.7)式，若观察到 $Y_1 = y_1$ 且 $Y_2 = y_2$，则区间(y_1, y_2)成为 θ 的置信系数为1/2的置信区间. 不过，这还可以做进一步分析.

既然两个观察值 X_1 和 X_2 都必须至少为 $\theta - (1/2)$，至多为 $\theta + 1/2$，我们可以确定 $y_1 \geqslant \theta - 1/2$ 且 $y_2 \leqslant \theta + 1/2$. 即，确信有：

$$
y_2 - (1/2) \leqslant \theta \leqslant y_1 + (1/2). \tag{7.5.8}
$$

现在我们假定$(y_2 - y_1) > 1/2$. 那么有 $y_1 < y_2 - (1/2)$，根据等式(7.5.8)可知 $y_1 < \theta$. 又因为 $y_1 + (1/2) < y_2$，同理由等式(7.5.8)可得 $\theta < y_2$. 这样若$(y_2 - y_1) > 1/2$，则 $y_1 < \theta < y_2$. 换言之，如果$(y_2 - y_1) > 1/2$，那么我们可以确定置信区间(y_1, y_2)包含了 θ 的未知值，尽管此区间的置信度只有1/2.

事实上，即使$(y_2 - y_1) \leqslant 1/2$，只要当 $y_2 - y_1$ 的值越接近1/2，我们就越能确认区间(y_1, y_2)包含 θ. 而且 $y_2 - y_1$ 的值越接近0，我们就越确认区间(y_1, y_2)不含 θ. 不过置信系数必须保持为1/2，且不依赖于 y_1 和 y_2 的观察值.

该例子也有助于解释7.1节最后的注记. 在该问题中，很自然地会用 $\overline{X}_2 = 0.5(X_1 + X_2)$来估计 θ. 用3.9节中的方法，我们可以得到 $\overline{X}_2$ 的概率密度函数：

$$
g(x) = \begin{cases} 4x - 4\theta + 2 & \text{若 } \theta - \dfrac{1}{2} < x \leqslant \theta, \\ 4\theta - 4x + 2 & \text{若 } \theta < x < \theta + \dfrac{1}{2}, \\ 0 & \text{其他.} \end{cases}
$$

图7-3表明概率密度函数 $g(x)$ 的图像为三角形，这样很简单地就可计算出 $\overline{X}_2$ 和 θ 接近的概率：

$$
\Pr\left(|\overline{X}_2 - \theta| < c\right) = 4c(1 - c), \quad 0 < c < 1/2.
$$

图7-3 例7.5.4中 $\overline{X}_2$ 的密度

举个例子，若 $c=0.3$，则 $\Pr(|\overline{X}_2-\theta|<0.3)=0.84$. 但是随机变量 $Z=Y_2-Y_1$ 包含了由计算所不能解释的有用的信息. 实际上，在给定 $Z=z$ 的条件下 $\overline{X}_2$ 的条件分布

413

是区间 $\left[\theta-\frac{1}{2}(1-z),\ \theta+\frac{1}{2}(1-z)\right]$ 上的均匀分布. 我们看到，观察值 z 越大，$\overline{X}_2$ 的可能取值的范围就越小. 特别地，给定 Z，$\overline{X}_2$ 接近于 θ 的条件概率为

$$\Pr(|\overline{X}_2-\theta|<c\,|Z=z)=\begin{cases}\dfrac{2c}{1-z} & \text{对于 } c\leqslant(1-z)/2,\\ 1 & \text{对于 } c>(1-z)/2.\end{cases}\tag{7.5.9}$$

例如，若 $z=0.1$，那么 $\Pr(|\overline{X}_2-\theta|<0.3\,|Z=0.1)=0.6667$，该值远远小于边缘概率 0.84. 由此说明仅仅由于估计的抽样分布相接近的概率很高而推断我们的估计十分接近于参数是不可靠的. 有可能其他信息告诉我们估计的接近程度并不如抽样分布那么好，或者是比所提出的抽样分布还要接近 [读者应该从另一种角度计算 $\Pr(|\overline{X}_2-\theta|<0.3\,|Z=0.9)$]. ◀

在下一节中，我们将要讨论用贝叶斯方法来分析取自均值 μ 和方差 σ^2 都未知的正态总体的随机样本. 我们先给 μ 和 σ^2 指定一个联合先验分布，再计算 μ 属于任意给定区间 (a,b) 的后验概率，可以证明[例如见 Degroot(1970)]，如果 μ 和 σ^2 的联合先验概率密度函数相当平滑，且属于 μ 和 σ^2 任意一个小集合的概率不大，而且如果样本容量 n 很大，那么均值 μ 属于一个特定置信区间 (a,b) 的置信度会与 μ 落入区间 (a,b) 的后验概率近似相等. 在下一节中会给出该近似等式的例子. 因此，在这些条件下，在实际应用中，由置信区间方法所得的结论与先验概率方法所得到的结果之间的差异是很小的. 但是这些方法的解释是不同的.

414

7.5.4 小结

设 $X_1,\cdots,X_n$ 是一个来自于均值为 μ、方差为 σ^2 的正态总体的独立随机样本，样本观察值为 $x_1,\cdots,x_n$. 令 $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ 和 $\sigma'^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$. 区间 $(\bar{x}_n-c\sigma'/n^{1/2},\bar{x}_n+c\sigma'/n^{1/2})$ 是 μ 在置信水平 γ 下的置信区间，其中 c 是自由度为 $n-1$ 的 t 分布的 $(1+\gamma)/2$ 分位点的值.

习题

1. 设 $X_1,\cdots,X_n$ 是来自于均值 μ 未知、方差 σ^2 已知正态总体的一个随机样本，令 Φ 为标准正态分布的分布函数，Φ^{-1} 为其逆函数. 证明：

$$\left(\bar{x}_n-\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{n^{1/2}},\bar{x}_n+\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{n^{1/2}}\right).$$

是 μ 在置信水平为 γ 下的置信区间，其中 $\bar{x}_n$ 是观察值的均值.

2. 设一个含有 8 个观察值的随机样本，它们来自于均值 μ 和方差 σ^2 均未知的正态总体. 其观察值分别是 3.1，3.5，2.6，3.4，3.8，3.0，2.9，2.2. 求 μ 在置信度分别为 (a) 90%，(b) 95%，(c) 99% 条件下的置信区间.

3. 设 $X_1,\cdots,X_n$ 是来自于均值 μ 和方差 σ^2 均未知的正态总体的一个随机样本. 令随机变量 L 表示可以由样本观察值来构造的 μ 的最小置信区间的长度，求在下述样本容量 n 和置信度 γ 条件下 $\mathrm{E}(L^2)$ 的值.

   a. $n=5$，$\gamma=0.95$.　b. $n=10$，$\gamma=0.95$.　c. $n=30$，$\gamma=0.95$.

   d. $n=8$，$\gamma=0.90$.　e. $n=8$，$\gamma=0.95$.　f. $n=8$，$\gamma=0.99$.

4. 设 $X_1,\cdots,X_n$ 是一个服从正态分布的随机样本，其中均值 μ 未知，方差 σ^2 已知. 求样本容量取多大时，才能存在一个置信区间使 μ 在置信度为 0.95 条件下置信区间的长度小于 0.01σ?

5. 设 $X_1,\cdots,X_n$ 是来自于一个均值 μ、方差 σ^2 均未知的正态分布的随机样本. 描述在已知置信度为 γ $(0<\gamma<1)$ 条件下构造 σ^2 的置信区间的方法. 提示：确定使得

$$\Pr\left[c_1<\frac{\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}{\sigma^2}<c_2\right]=\gamma$$ 成立的常数 c_1 和 c_2.

6. 设 $X_1,\cdots,X_n$ 是来自于均值 μ 未知的指数分布的随机样本. 描述一个在指定置信度 γ $(0<\gamma<1)$ 条件下，构造 μ 的置信区间的方法. 提示：确定常数 c_1 和 c_2，使得

$$\Pr\left[c_1<(1/\mu)\sum_{i=1}^{n}X_i<c_2\right]=\gamma.$$

7. 在《消费报告》杂志 1986 年 6 月期中，给出了一些牛肉热狗卡路里含量的数据. 以下是 20 种不同品牌热狗的卡路里含量：

   186，181，176，149，184，190，158，139，175，148，
   152，111，141，153，190，157，131，149，135，132.

   假设它们是均值 μ 和方差 σ^2 均未知的正态型随机变量的 20 个相互独立的观察值. 求卡路里均值 μ 的置信度为 90% 的置信区间.

415

8. 在例 7.5.4 的末尾，在已知 $Z=0.9$ 条件下，计算 $|\bar{X}_2-\theta|<0.3$ 的概率. 解释为什么概率会这么大?

9. 在例 7.5.4 中，假设我们观察到 $X_1=4.7$，$X_2=5.3$.

   a. 同例 7.5.4 中条件，求置信度为 50% 的置信区间.
   b. 求与观察数据一致的 θ 可能取值的区间.
   c. 50% 的置信区间与 θ 的可能取值区间相比是大还是小?
   d. 计算例 7.5.4 中随机变量 $Z=Y_2-Y_1$ 的值.
   e. 利用公式(7.5.9)，计算给定 Z 下 $|\bar{X}_2-\theta|<0.1$ 的条件概率，其中 Z 等于 (d) 中所计算出来的值.

10. 在习题 9 中，假设为 θ 指定一个概率密度函数为 $\xi(\theta)=0.1\exp(-0.1\theta)$（其中 $\theta>0$）的先验分布（这是参数为 0.1 的指数分布）.

a. 证明：习题9中在给定观察数据条件下的后验概率密度函数为

$$\xi(\theta|x)=\begin{cases}4.122\exp(-0.1\theta) & \text{对于 } 4.8<\theta<5.2,\\ 0 & \text{其他.}\end{cases}$$

b. 计算 $|\theta-\bar{x}_2|<0.1$ 的后验概率，其中 $\bar{x}_2$ 是观察数据的均值.

c. 计算 θ 落入习题9（a）中所得的置信区间的概率.

d. 你能否解释为什么习题9中（b）的结果与（e）的结果这么接近？

提示：比较（a）中的后验概率密度函数与公式(7.5.9)中的函数.

416

7.6 无偏估计

令 δ 是参数 θ 的函数 g 的估计量. 如果对于任意 θ 有 $E_\theta[\delta(\boldsymbol{X})]=g(\theta)$ 成立，则我们称 δ 是无偏的. 在本节中将给出几个无偏估计的例子.

427

7.6.1 无偏估计的定义

再考虑这种情况，变量 $X_1,\cdots,X_n$ 来自于含有未知参数 θ 的分布的随机样本. 在这类问题中我们希望可以找到一个估计量 $\delta(X_1,\cdots,X_n)$ 以高概率接近于 θ. 换句话说，我们想得到一个估计量 δ，它随 θ 这样地变化：无论 θ 的真值是多少，δ 的概率分布总是集中于 θ 的真值附近.

例如，假设变量 $X_1,\cdots,X_n$ 是来自于均值 θ 未知、方差为1的正态分布的一个随机样本，这时样本均值 $\overline{X}_n$ 是 θ 的极大似然估计(M. L. E). 估计量 $\overline{X}_n$ 服从均值为 θ、方差为 $1/n$ 的正态分布，所以有理由认为它是 θ 的一个好的估计量. 无论 θ 的值多大或多小，该分布都集中于未知参数 θ 附近.

这种想法可导致如下定义：如果对于 θ 的任意值，有 $E_\theta[\delta(X_1,\cdots,X_n)]=\theta$ 成立，则称估计量 $\delta(X_1,\cdots,X_n)$ 为参数 θ 的*无偏估计*. 若一个估计量不是无偏的，则称其为*有偏估计*. 估计值的数学期望与被估计的参数 θ 之间的差异称为估计值的偏差. 即，θ 的一个估计量 δ 的偏差 $E_\theta[\delta(X_1,\cdots,X_n)]-\theta$，当且仅当偏差对于任意 θ 都为0时才称 δ 是无偏的.

在一个均值 θ 未知的正态样本中，$\overline{X}_n$ 是 θ 的无偏估计，这是因为对于 $-\infty<\theta<\infty$ 都有 $E_\theta(\overline{X}_n)=\theta$ 成立.

如果某个参数 θ 的一个估计量是无偏的，那么 δ 的分布必定会随 θ 的值而改变，这是由于该分布的均值是 θ. 然而需要强调的是，这样的分布有可能会集中于 θ 附近也可能会分散得很开. 比如，等可能地被高估或低估100万单元的估计量是一个无偏估计量，但它不会产生一个接近于 θ 的估计. 因此，仅估计量的无偏性这一事实不能说明该估计量是一个好的估计量或者说它是合理的. 不过，如果同时 θ 的一个无偏估计有很小的方差，则这个估计量的分布将必然集中于均值 θ 附近，即该估计量接近于 θ 的概率很大.

由于上述这些原因，一个特定问题中无偏估计量的研究就是投入大量精力去寻找一个具有较小方差的无偏估计量. 如果估计量 δ 是无偏的，那么它的均方误差(M. S. E) $E_\theta[(\delta-\theta)^2]$就等于它的方差 $Var_\theta(\delta)$. 因此，寻找一个较小方差的估计量也就等价于寻找一个有较小均方误差的无偏估计量.

θ 的函数的无偏估计

很容易将无偏估计量的定义推广到对于 θ 的任意函数的估计中去. 例如，令 $X_1,\cdots,X_n$是来自于均值μ 和方差 σ^2 均未知的正态总体的一个随机样本，令 $\theta=(\mu,\sigma^2)$为参数. 我想要估计 $g(\theta)=\mu$. 一般地，如果对任意的 θ，$E_\theta[\delta(X_1,\cdots,X_n)]=g(\theta)$成立，则称 $\delta(X_1,\cdots,X_n)$为 $g(\theta)$ 的一个*无偏估计量*. 在正态分布的情况中，我们很容易证明 $\overline{X}_n$ 是 μ 的一个无偏估计量. 更普遍地，如果 $X_1,\cdots,X_n$ 是来自于参数为 θ、均值 μ 是有限的任意一个分布的样本，那么 $\mu=E_\theta(X_i)$ 是 θ 的一个函数. 由于等式$E(\overline{X}_n)=\mu$ 恒成立，因此很容易证明均值 $\overline{X}_n$ 为 μ 的一个无偏估计. 428

7.6.2　方差的无偏估计

1. 来自任意分布的样本

我们已经注意到如果 $X_1,\cdots,X_n$ 是一个来自于均值 μ 未知的任意的随机样本，那么该样本的均值 $\overline{X}_n$ 是 μ 的一个无偏估计. 现在我们假设下面分布中的方差 σ^2 也是未知的，并且将确定 σ^2 的一个无偏估计.

既然样本均值是均值 μ 的一个无偏估计，那么很自然地会首先考虑样本的方差 $\hat{\sigma}_0^2=(1/n)\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$，并且确定 $\hat{\sigma}_0^2$ 是否是方差 σ^2 的一个无偏估计，我们运用恒等式

$$\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}(X_i-\overline{X}_n)^2+n(\overline{X}_n-\mu)^2.$$

因此可以得到以下结果

$$\begin{aligned}E(\hat{\sigma}_0^2)&=E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\right]\\&=E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]-E[(\overline{X}_n-\mu)^2].\end{aligned}\tag{7.6.1}$$

既然每一个观察值 X_i 有均值 μ 和方差 σ^2，并且对于 $i=1,\cdots,n$,有 $E[(X_i-\mu)^2]=\sigma^2$. 因此，

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]=\frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2]=\frac{1}{n}n\sigma^2=\sigma^2.\tag{7.6.2}$$

此外，样本均值 $\overline{X}_n$ 的均值为 μ 和方差为 σ^2/n. 因此，

$$E[(\overline{X}_n-\mu)^2]=Var(\overline{X}_n)=\frac{\sigma^2}{n}.\tag{7.6.3}$$

由等式(7.6.1)，(7.6.2)和(7.6.3)有

$$E(\hat{\sigma}_0^2)=\sigma^2-\frac{1}{n}\sigma^2=\frac{n-1}{n}\sigma^2. \tag{7.6.4}$$

从等式(7.6.4)可以看出样本方差 $\hat{\sigma}_0^2$ 不是 σ^2 的无偏估计，这是因为它的期望值是$[(n-1)/n]\sigma^2$，而不是 σ^2. 但是，如果 $\hat{\sigma}_0^2$ 乘以因子 $n/(n-1)$就可得到一个统计量 $\hat{\sigma}_1^2$，而 $\hat{\sigma}_1^2$ 的期望值的确是 σ^2. 因此，$\hat{\sigma}_1^2$ 是 σ^2 的无偏估计，从而我们有以下结论：

如果随机变量 $X_1,\cdots,X_n$ 是一组来自于任意分布的随机样本，并且它的方差 σ^2 是未知的，那么估计

429

$$\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$$

是 σ^2 的无偏估计.

换言之，虽然 $\hat{\sigma}_0^2$ 不是 σ^2 的无偏估计，但是由离差平方和 $\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 除以 $n-1$，而不是 n 所得到的估计量 $\hat{\sigma}_1^2$ 总是 σ^2 的无偏估计. 正是由于这个原因，在许多教科书中，样本方差最初就被定义为 $\hat{\sigma}_1^2$，而不是 $\hat{\sigma}_0^2$.

注： 正态随机样本的特例. 当 $X_1,\cdots,X_n$ 来自于均值为 μ、方差为 σ^2 的正态分布时，估计量 $\hat{\sigma}_0^2$ 等于其方差 σ^2 的极大似然估计 $\hat{\sigma}^2$. 同时，$\hat{\sigma}_1^2$ 与出现在 μ 的置信区间内的随机变量 σ'^2是相同的. 由于在本节中我们所讨论的是一般分布，对它来说 σ^2 有可能是某一函数 $g(\theta)$，且其极大似然估计与 $\hat{\sigma}_0^2$ 完全不同. 因此，对于本节中的这些估计量选用了不同的名字(可见习题1中有一个这样的例子).

2. 来自某个特定分布族的样本

当我们假设 $X_1,\cdots,X_n$ 是一组来自于某个特定分布族的随机样本，例如泊松分布，我们一般不仅仅考虑 $\hat{\sigma}_1^2$，而且还考虑方差的其他无偏估计. 例如，假设该样本实际上是来自于均值为 θ 未知的泊松分布. 我们已经知道 $\overline{X}_n$ 是均值 θ 的无偏估计. 此外，由于泊松分布的方差也等于 θ，这就表示 $\overline{X}_n$ 也是方差的无偏估计. 因此，在本例中，$\overline{X}_n$ 和 $\hat{\sigma}_1^2$ 都是未知方差 θ 的无偏估计. 再者，任意满足 $\alpha\overline{X}_n+(1-\alpha)\hat{\sigma}_1^2$ 形式的 $\overline{X}_n$ 与 $\hat{\sigma}_1^2$ 组合，也都是 θ 的无偏估计，这里 α 是一个给定常数($-\infty<\alpha<\infty$)，这是因为它的期望值

$$E\left[\alpha\overline{X}_n+(1-\alpha)\hat{\sigma}_1^2\right]=\alpha E(\overline{X}_n)+(1-\alpha)E(\hat{\sigma}_1^2)=\alpha\theta+(1-\alpha)\theta=\theta.$$

θ 其他的无偏估计量也可以用类似方法构造.

如果要选用一个无偏估计量，那么问题就在于确定哪一个无偏估计量有最小的方差或者有最小的均方误差. 我们先不立即解决这一问题. 但由相关知识可知对于 θ 所有可能的取值，在所有 θ 的无偏估计中，估计量 $\overline{X}_n$ 的方差最小. 这一结果并不意外. 我们可以知道 $\overline{X}_n$ 是 θ 的充分统计量，并且在某些方面我们可以把注意力仅集中在充分统计量的函数上(也可见本节末尾中的习题13).

7.6.3 正态分布的方差的估计

假设 $X_1,\cdots,X_n$ 是来自于均值 μ 和方差 σ^2 都未知的正态分布的一个随机样本，我们将考虑 σ^2 估计的问题. 从这部分的讨论中，可以知道估计量 $\hat{\sigma}_1^2$ 是 σ^2 的一个无偏估计量. 此外，从6.4节中我们知道样本方差 $\hat{\sigma}_0^2$ 是 σ^2 的极大似然估计. 我们想要确定估计量 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_1^2$ 哪一个的均方误差 $\mathrm{E}[(\hat{\sigma}_i^2-\sigma^2)^2]$ 要更小些，而且是否有 σ^2 其他的估计量比 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_1^2$ 的均方误差都要小.

430

估计量 $\hat{\sigma}_0^2$ 和估计量 $\hat{\sigma}_1^2$ 都具有以下的形式：

$$T^2=c\sum_{i=1}^{n}(X_i-\overline{X}_n)^2, \tag{7.6.5}$$

其中对于 $\hat{\sigma}_0^2$ 来说 $c=1/n$，而对 $\hat{\sigma}_1^2$ 来说 $c=1/(n-1)$. 我们现在可以确定任意一个具有等式(7.6.5)形式的估计量的均方误差，然后确定均方误差最小时 c 的值. 我们将要证明一个有趣的性质，相同的 c 值可使参数 μ 和 σ^2 对于所有可能取值都使均方误差最小化. 因此，在所有具有等式(7.6.5)形式的估计量中，只有一个估计量会含有参数 μ 和 σ^2 对于所有可能的取值都使均方误差最小化.

在7.3节中已经证明了当 $X_1,\cdots,X_n$ 是来自于正态分布的一个随机样本，随机变量 $\sum_{i=1}^{n}(X_i-\overline{X}_n)^2/\sigma^2$ 服从自由度为 $n-1$ 的 χ^2 分布. 由等式(7.2.2)知变量的均值是 $n-1$，方差为 $2(n-1)$. 因此，如果 T^2 如等式(7.6.5)所定义的这样，则

$$\mathrm{E}(T^2)=(n-1)c\sigma^2,\mathrm{Var}(T^2)=2(n-1)c^2\sigma^4. \tag{7.6.6}$$

这样，通过4.3节的习题5，T^2 的均方误差可以表示如下：

$$\begin{aligned}\mathrm{E}[(T^2-\sigma^2)^2]&=[\mathrm{E}(T^2)-\sigma^2]^2+\mathrm{Var}(T^2)\\&=[(n-1)c-1]^2\sigma^4+2(n-1)c^2\sigma^4\\&=[(n^2-1)c^2-2(n-1)c+1]\sigma^4.\end{aligned} \tag{7.6.7}$$

在等式(7.6.7)中 σ^4 的系数是 c 的一个简单二次函数. 因此，对于 σ^4 的每一个值，c 的最小值可以通过基本的微分来求出，得 $c=1/(n+1)$.

总的来说，我们可以证明了如下的事实：在所有具有等式(7.6.5)形式的 σ^2 的估计量中，对于所有 μ 和 σ^2 的可能取值都有最小均方误差的估计量是 $T_0^2=[1/(n+1)]\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$. 特别地，$T_0^2$ 具有比极大似然估计 $\hat{\sigma}_0^2$ 和无偏估计量 $\hat{\sigma}_1^2$ 都要小的均方误差. 因此，估计量 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_1^2$，还有所有其他的具有等式(7.6.5)形式的，但 $c\neq 1/(n+1)$ 的估计量都是不允许的. 进而 C. Stein 在1964年证明的，当估计量 T_0^2 被其他估计量支配时，那么它本身也是不允许的.

在本节最后的习题6中比较了估计量 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_1^2$. 当然，当样本容量 n 很大时，无论是取 n，$n-1$ 还是 $n+1$ 作为除数来估计 σ^2，区别都是很小的，这三个估计量 $\hat{\sigma}_0^2$，$\hat{\sigma}_1^2$ 和 T_0^2 也是近乎相等的.

7.6.4 扩展研究：无偏估计的局限性

无偏估计的概念在统计学的发展中起了十分重要的作用. 无偏估计比有偏估计要好的看法在目前的统计实践中更加普遍. 其实，科学家想要得到的结果是有偏的还是被假设成有偏的？无偏估计的专门术语的理论使无偏估计量的使用看上去更令人满意. 431

然而，正如本节内容中的解释一样，一个无偏估计量的好坏一定是用它的方差或是它的均方误差来评判. 我们已经证明了当要估计一个正态分布的方差时，或在其他的问题中，总会存在一个有偏估计量，对于参数的每个可能取值都要比每个无偏估计量有更小的均方误差. 此外，还可以证明一个贝叶斯估计量，尽管已经利用了所有与参数有关的先验信息，并使总的均方误差最小化，但只有在参数可以完美地被估计出来的问题中才是无偏的.

下面我们来介绍无偏估计理论的其他一些局限性.

1. 无偏估计不存在

在许多问题中参数的无偏估计或者是某个必须要估计的参数的特定函数的无偏估计量是不存在的. 例如，$X_1,\cdots,X_n$ 服从参数为 p 的 n 重伯努利试验，其中 p 未知 $(0\leqslant p\leqslant 1)$，则样本均值 $\overline{X}_n$ 是 p 的无偏估计，但是可以证明 $p^{1/2}$ 的无偏估计不存在(详见习题7). 而且如果已知例子中 p 必须落在区间 $1/3\leqslant p\leqslant 2/3$ 内，那么对于 p 可能取值的区间而言，不存在 p 的无偏估计.

2. 不适当的无偏估计

考虑参数 p 未知的伯努利试验中的一个无穷序列 $(0<p<1)$，令 X 表示在第一次成功前失败的次数. 那么 X 服从一个几何分布，其概率函数如等式(5.5.3)中所定义. 如果想要从观察值 X 来估计 p 值，可以得到(详见习题8)：当 $X=0$ 时，p 的无偏估计为1，而当 $X>0$ 时，p 的无偏估计为0. 该估计看上去不太恰当. 例如：假设在第二次试验中取得了第一次成功，即，如果 $X=1$，概率 p 为0是错误的. 同样地，如果 $X=0$(第一次试验就成功)，估计 p 接近于1也是错误的.

再举一个不恰当无偏估计的例子. 假设随机变量 X 服从均值为 $\lambda(\lambda>0)$ 且未知的泊松分布，要求估计 $\mathrm{e}^{-2\lambda}$ 的值. 我们可以看出(详见习题9)：当 X 是偶数时，$\mathrm{e}^{-2\lambda}$ 无偏估计为1；当 X 为奇数时，$\mathrm{e}^{-2\lambda}$ 无偏估计为 -1. 有两个原因表明该估计不太恰当：第一，参数 $\mathrm{e}^{-2\lambda}$ 必须在 $(0,1)$ 的范围内，由此得出参数 $\mathrm{e}^{-2\lambda}$ 的估计是1或 -1；第二，估计值只依赖于 X 的奇偶性，而不是 X 的大小.

3. 忽略信息

对于无偏估计的概念最后要批判的是由于总是运用参数 θ 无偏估计的准则(如果这样的无偏估计存在的话)，有时会忽略掉一些已有的有用的信息. 例如：假设在某个电路中，平均电压 θ 未知，用电压计来测电压，我们发现读数 X 服从正态分布，其均值 θ、方差 σ^2 为已知，看到电压计上的观察值为205伏特. 在本例中，X 就是 θ 432

的无偏估计. 一个希望用无偏估计的科学家就会把 205 伏特作为 θ 的估计值.

但是，假设该科学家把 205 伏特作为 θ 的估计值后，他会发现在他读数时，伏特表并没有正常工作. 用伏特表测 280 伏特以下的值时，伏特表读数是精确的，但当电压大于 280 伏特，测出来的就不精确了. 既然实际读出的值是 205 伏特，该值并未受到伏特表缺陷的影响. 不过，当得到的读数 X 的分布不是均值为 θ 的正态分布时，该观察值也就不是 θ 的无偏估计了. 因此，如果该科学家依旧想要用无偏估计，他就必须把 θ 的估计值从 205 伏特改为其他的值.

忽略观察值是精确的这一事实看起来不太容易被接受. 由于实际观察到的读数为 205 伏特，它应该和仪表没有故障时的观察值相同. 既然观测的读数正确，看起来已经有错误读数的事实和 θ 的估计是不相关. 但是，该可能性尽管我们知道它不会发生，但它确实改变了 X 的样本空间及 X 的概率分布，它也会改变 θ 的无偏估计形式.

7.6.5 小结

如果对于 θ 所有可能的取值都有 $E_\theta[\delta(X_1,\cdots,X_n)]=g(\theta)$，则称 $\delta(X_1,\cdots,X_n)$ 为 $g(\theta)$ 的无偏估计. 一个无偏估计的均方误差等于该估计的方差.

习题

1. 设 $X_1,\cdots,X_n$ 是来自于均值为 θ 的泊松分布的一个随机样本.
   a. 将 $\mathrm{Var}_\theta(X_i)$ 表示为 $\sigma^2=g(\theta)$ 的函数形式.
   b. 求 $g(\theta)$ 的均方误差，并证明它是无偏的.
2. 假设 X 是一个分布未知的随机变量，但是知道它的所有阶矩 $E(X^k)$ 是有限的. 假设 $X_1,\cdots,X_n$ 是来自于该分布的一个随机样本. 证明：当 $k=1,2,\cdots$，$(1/n)\sum_{i=1}^n X_i^k$ 是 $E(X^k)$ 的无偏估计.
3. 条件同第 2 题，求 $[E(X)]^2$ 的无偏估计. 提示：$[E(X)]^2=E(X^2)-\mathrm{Var}(X)$.
4. 假设一个随机变量 X 服从参数 $p(0<p<1)$ 未知的几何分布（几何分布同 5.5 节中的定义）. 求 $1/p$ 的无偏估计统计量 $\delta(X)$.
5. 假设随机变量 X 服从均值为 λ 的泊松分布，其中 λ 未知($\lambda>0$). 求 e^λ 的无偏估计统计量 $\delta(X)$. 提示：如果 $E[\delta(X)]=e^\lambda$，那么
$$\sum_{x=0}^{n}\frac{\delta(x)e^{-\lambda}\lambda^x}{x!}=e^\lambda.$$
等式两边同乘 e^λ，右边以 λ 幂序列的形式展开，然后对于 $x=0,1,2,\cdots$，从等式两边计算 λ^x 的系数.

433

6. 假设 $X_1,\cdots,X_n$ 是来自于正态分布的一个随机样本，其均值 μ 和方差 σ^2 都未知. 令 $\hat{\sigma}_0^2$ 和 $\hat{\sigma}_1^2$ 是 σ^2 的两个估计量，其定义如下：
$$\hat{\sigma}_0^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X}_n)^2,\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X}_n)^2.$$
证明：对于 μ 和 σ^2 的所有可能取值，$\hat{\sigma}_0^2$ 的均方误差要比 $\hat{\sigma}_1^2$ 的均方误差小.

7. 假设 $X_1,\cdots,X_n$ 是来自于参数 p 未知的 n 重伯努利试验($0\leqslant p\leqslant 1$). 证明：每个函数 $\delta(X_1,\cdots,X_n)$ 的期望是关于 p 的一个次数不超过 n 的多项式.
8. 假设一个随机变量 X 服从几何分布，其参数 p 未知($0<p<1$). 证明：p 的唯一的无偏估计量 $\delta(X)$ 满足：$\delta(0)=1$，而当 $X>0$ 时有 $\delta(X)=0$.
9. 假设一个随机变量 X 服从泊松分布，其均值 λ 未知($\lambda>0$). 证明 $e^{2\lambda}$ 的唯一的无偏估计量 $\delta(X)$ 满足：假如 X 是偶数时，$\delta(X)=1$，而当 X 是奇数时，$\delta(X)=-1$.
10. 考虑一个伯努利试验的无限序列，其中参数 p 未知($0<p<1$)，假设连续取样直到恰好得到 k 次成功，这里 k 为固定的整数值($k\geqslant 2$). 定义 N 是为获得 k 次成功数所需要的总的试验次数. 证明：估计量 $(k-1)/(N-1)$ 是 p 的无偏估计量.
11. 假定一种特定的药物提供给两种不同类型的动物 A 和 B. 已知 A 类型动物的平均反应跟 B 类型的一样，但是该均值的共同值 θ 未知并且必须要估计. 此外还知 A 类动物反应的方差是 B 类动物反应的方差的 4 倍. 令 $X_1,\cdots,X_m$ 表示 m 个 A 类动物反应的一个随机样本，将 $Y_1,\cdots,Y_n$ 定义为 n 个 B 类动物的一个独立的随机样本. 最后考虑估计量 $\hat{\theta}=\alpha\overline{X}_m+(1-\alpha)\overline{Y}_n$.

   a. 求当 α，m 和 n 为何值时 $\hat{\theta}$ 是 θ 的一个无偏估计?

   b. 求当 m 和 n 值固定时，α 取何值时，无偏估计量具有最小方差?
12. 假定某总体是由 k 个不同的阶层的个体组成的($k\geqslant 2$)，在总体中，个体属于阶层 i 的比例为 p_i,($i=1,\cdots,k$)，且 $p_i>0$，$\sum_{i=1}^{k}p_i=1$. 我们对于估计总体中某个特征的均值 μ 感兴趣. 在第 i 个阶层的个体中，该特征的均值为 μ_i 且方差为 σ_i^2，其中 μ_i 值未知而 σ_i^2 值已知. 假定从总体中按如下方法抽取一个分层样本：在每一个阶层 i 中，取 n_i 个个体的随机样本并测出每一个个体特征. 从 k 个阶层中所取的样本是互相独立的. 令 $\overline{X}_i$ 表示阶层 i 样本中 n_i 个测量值的平均值.

   a. 证明：$\mu=\sum_{i=1}^{k}p_i\mu_i$，且 $\hat{\mu}=\sum_{i=1}^{k}p_i\overline{X}_i$ 是 μ 的无偏估计量.

   b. 令 $n=\sum_{i=1}^{k}n_i$ 表示 k 个样本的总观察值数. 当 n 固定时，找出 $n_1,\cdots,n_k$ 的值使得 $\hat{\mu}$ 的方差最小.
13. 假定 $X_1,\cdots,X_n$ 来自于概率密度函数或者概率函数为 $f(x|\theta)$ 的分布的一个随机样本. 其中参数 θ 值未知. 令 $\boldsymbol{X}=(X_1,\cdots,X_n)$，$T$ 为一个统计量. 假定 $\delta(\boldsymbol{X})$ 是 θ 的一个无偏估计量，且使得 $E_\theta[\delta(\boldsymbol{X})|T]$ 不依赖于 θ.（假如 T 是一个充分统计量，那么对于每个估计量 δ 都是成立的，该条件在其他例子中也成立). 令 $\delta_0(T)$ 表示在给出的 T 条件下 $\delta(X)$ 的条件均值.

   a. 证明：$\delta_0(T)$ 也是 θ 的一个无偏估计量.

   b. 证明：对于 θ 的每个可能值，都有 $\mathrm{Var}_\theta(\delta_0)\leqslant\mathrm{Var}_\theta(\delta)$.
14. 假定 $X_1,\cdots,X_n$ 是来自于区间 $[0,\theta]$ 上均匀分布的一个随机样本，参数 θ 的值未知，令 $Y=\max(X_1,\cdots,X_n)$. 证明：$[(n+1)/n]Y_n$ 是 θ 的无偏估计量.
15. 假定随机变量 X 只可以取 5 个值，$x=1,2,3,4,5$，其相应概率如下：

$$f(1|\theta)=\theta^3,\quad f(2|\theta)=\theta^2(1-\theta),$$
$$f(3|\theta)=2\theta(1-\theta),\quad f(4|\theta)=\theta(1-\theta)^2,\quad f(5|\theta)=(1-\theta)^3.$$

434

其中参数 θ 的值未知 $(0\leqslant\theta\leqslant1)$.

a. 证明：对于每一个 θ 值，所给的五个概率值的和为 1.

b. 考虑一个估计量 $\delta_c(X)$ 具有如下的形式：

$$\delta_c(1)=1,\delta_c(2)=2-2c,\delta_c(3)=c,\delta_c(4)=1-2c,\delta_c(5)=0.$$

试证明：对于每个常数 c，$\delta_c(X)$ 是 θ 的一个无偏估计量.

c. 令 θ_0 是一个满足 $0<\theta_0<1$ 的值. 确定常数 c 使得当 $\theta=\theta_0$ 时，对于所有其他 c 的值，都有 $\delta_{c_0}(X)$ 的方差小于 $\delta_c(X)$ 的方差.

16. 再次考虑习题 3 中的条件. 假定 $n=2$，而且我们观察到 $X_1=2$，$X_2=-1$. 计算习题 3 中 $[\mathrm{E}(X)]^2$ 的无偏估计量的值. 描述你所发现的该估计量的缺点.

435

第8章 假设检验

8.1 假设检验问题

我们考虑这样一个问题，确定在临床试验中成功的概率 P 是否大于某个值 p_0，即假设检验问题. 一般来说，假设检验所关心的是确定参数 θ 是属于参数空间中的一个子集还是它的补集. 当 θ 是一维的，通常该两个子集中至少有一个是区间，但可能是退化的. 在本节中，我们引入有关假设检验的符号和一些常用的计算方法. 同时还会说明假设检验和置信区间的等价关系.

8.1.1 原假设和备择假设

在本章中，再一次考虑那些包含着一定取值于某一个参数空间 Ω 中的未知参数 θ 的统计问题. 现在假设 Ω 可以被分成两个互不相交的子集 Ω_0 和 Ω_1，那么统计学家要做的就是确定 θ 到底是属于 Ω_0 还是 Ω_1.

设 H_0 表示 $\theta\in\Omega_0$ 的假设，H_1 表示 $\theta\in\Omega_1$ 的假设. 由于集合 Ω_0 和 Ω_1 互不相交且 $\Omega_0\cup\Omega_1=\Omega$，显然假设 H_0 和 H_1 必然有一个是正确的. 统计学家必须要做的就是决定是接受 H_0 还是 H_1. 称该类只有2个可能需决定的问题，为假设检验的问题. 如果统计学家做了一个错误的决定，他可能遭受一定的损失或者支付一定的费用. 在许多问题上，他都有机会在做决定之前进行观察，然后观察到的值将会提供给他关于 θ 值的信息. 决定到底接受 H_0 的假设还是 H_1 的假设的过程被称为检验过程或者简单地称为检验.

449

讨论至今，我们一直必须把 H_0 假设和 H_1 假设平等地对待. 然而在大多数问题中，这两个假设是要区别对待的. 为了区分他们，称 H_0 为原假设，H_1 为备择假设. 在本章后面的部分以及本书剩余部分中所有有关假设检验的问题中，我们将会运用这种专用术语.

对于统计学家做决策的一种描述是他可以接受 H_0 或者 H_1. 然而，由于只有这两种可能的决定，所以接受 H_0 就等价于拒绝 H_1，同样地，接受 H_1 也就意味着拒绝 H_0. 我们将在对假设检验的讨论中运用所有这些描述. 一些作者用“不拒绝 H_0”来代替“接受 H_0”.

例 8.1.1 埃及人头盖骨 Manly(1986, p.4)报道了，在埃及发现的不同时期的人类头骨大小的测量结果. 这些数据是由 Thomson 和 Bandall-Maciver(1905)提供的. 有一个时间阶段大概是公元前4000多年，我们可能把宽度测量的数据（单位mm）看作服从均值为未知数 θ、方差为27的正态分布的随机变量. 问题是如何比

较 θ 和人类现在大概是 140mm 的头颅宽度．参数空间 Ω 可取为正数集，而且取 Ω_0 为 $(0,140]$，同时 $\Omega_1=(140,\infty)$．在该种情况下，可以把原假设和备择假设规定为：

$$H_0:\theta\leqslant 140,$$
$$H_1:\theta>140.$$

更具实际性地，可以假设宽度测量数据的均值和方差都是未知的．即，每一次测量都是均值为 μ、方差为 σ^2 的正态随机变量．在这种情况下，参数将会是二维的．例如 $\theta=(\mu,\sigma^2)$，参数空间 Ω 将是实数对的集合．在本问题中，$\Omega_0=(0,140]\times(0,\infty)$ 且 $\Omega_1=(140,\infty)\times(0,\infty)$．因为这里的假设仅仅关心第一坐标 μ，所以可以把需要检验的假设写成：

$$H_0:\mu\leqslant 140,$$
$$H_1:\mu>140.$$ ◀

8.1.2　简单假设和复合假设

假设 $X_1,\cdots,X_n$ 是服从概率密度函数或概率函数是 $f(x|\theta)$ 的分布的一个随机样本，这里参数 θ 的值必须落在参数空间 Ω 中，Ω_0 和 Ω_1 互不相交且 $\Omega_0\cup\Omega_1=\Omega$．现在需要检验的是以下的假设：

$$H_0:\theta\in\Omega_0,$$
$$H_1:\theta\in\Omega_1.$$

对于 $i=0$ 或者 $i=1$，集合 Ω_i 可能只包含一个 θ 值，因此说假设 H_i 是一个简单假设．如果集合 Ω_i 包含不仅一个 θ 值，那么就说 H_i 假设是一个复合假设．在一个简单假设下，观察值的分布是完全确定的．在一个复合假设前提下，我们只能确定观察值的分布是属于一个确定的组里．比如说，一个简单的原假设 H_0 必然有这样的形式：

450

$$H_0:\theta=\theta_0. \tag{8.1.1}$$

当 θ 是一维的，有两个比较常用的复合假设的形式．单边原假设是 $H_0:\theta\leqslant\theta_0$ 或者 $H_0:\theta\geqslant\theta_0$ 以及相应的单边备择假设是 $H_1:\theta>\theta_0$ 或者 $H_1:\theta<\theta_0$（见例 8.1.1）．若原假设是简单的，就像例 8.1.1 那样，那么备择假设通常是双边的，$H_1:\theta\neq\theta_0$（见例 8.1.2）．这些类型的假设将会在 8.3 节和 8.4 节中更加详细讨论．

8.1.3　临界域和检验统计量

现在考虑如下列形式的假设检验问题：

$$H_0:\theta\in\Omega_0,$$
$$H_1:\theta\in\Omega_1.$$

假定在统计学家决定接受哪个假设之前，她能够观察一个来自于涉及未知参数 θ 的分布的一个随机样本 $X_1,\cdots,X_n$．用 S 代表一个 n 维随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的样本

空间. 换句话说，S 是所有可能取到的随机样本结果的集合.

在这类问题中，统计人员通过划分样本空间 S 成两个子集来规定一个检验过程. 一个子集包含的 X 值将导致统计人员接受 H_0，而另一个子集包含的 X 值将使人们拒绝 H_0 因而接受 H_1. 导致拒绝原假设 H_0 的子集称为检验的临界域. 概括地说，检验过程是通过确定检验的临界域来决定的. 临界域的补集必须包含接受原假设 H_0 的所有样本点.

在大多数假设检验问题中，临界域是通过检验统计量 $T=r(X)$ 来定义的，它是数据 X 的一个函数. 典型地，使用统计量 T 的一个检验通过观测统计量 T 是落入一些固定的区域还是在一些固定的区域之外来拒绝原假设. 例如，如果在 $T\geqslant c$ 时，拒绝 H_0，那么临界域就是 $r(x)\geqslant c$ 的所有 x 的集合. 一旦找到了统计量，较简单的办法是集中于统计量 T 是落在固定区域内还是外，而忽略临界域. 这本书接下来的所有检验都建立在统计量的基础上. 大多数的检验能写成"如果 $T\geqslant c$，则拒绝 H_0"的形式（例 8.1.3 是一个极少的例外）.

例 8.1.2 已知方差的正态总体均值的检验 假定 $X=(X_1,\cdots,X_n)$ 是取自于一个未知均值 μ、已知方差 σ^2 的正态总体的一个随机样本. 设检验假设：

$$\begin{aligned} H_0&:\mu=\mu_0,\\ H_1&:\mu\neq\mu_0. \end{aligned} \tag{8.1.2}$$

451

如果样本均值 $\overline{X}_n$ 远离 μ_0，则看上去可以合乎情理地拒绝原假设 H_0. 在该种情况下，定义 $T=|\overline{X}_n-\mu_0|$，并且对于每一个 $c>0$，创建一个检验过程 δ_c，如果统计量 T 的值超过了给定的 c 的值，即 $T\geqslant c$，则拒绝原假设 H_0. ◀

可以如例 8.1.2 通过直觉判断来选择检验统计量，也可以基于理论上的考虑来选择. 在 8.2~8.4 节中将对于涉及单个参数的各种问题中如何选择检验统计量来做理论探讨. 虽然这些理论上的结论在应用它们的情形下可以提供最优的检验，但是许多实际问题并不满足应用这些结果所要求的条件.

8.1.4 功效函数和错误的类型

检验过程的特点可以描述为，对于每一个属于 Ω 的 θ，给出拒绝原假设的概率 $\pi(\theta|\delta)$，或给出接受原假设的概率 $1-\pi(\theta|\delta)$. 称函数 $\pi(\theta|\delta)$ 为检验的*功效函数*. 这样，如果用 C 代表检验 δ 的临界域，则功效函数被

$$\pi(\theta|\delta)=\Pr(X\in C|\theta),\text{对 }\theta\in\Omega. \tag{8.1.3}$$

这一关系所限定.

由于功效函数对于参数 θ 的每一个可能的取值，指出了 δ 将拒绝原假设 H_0 的概率，那么理想化的功效函数应当满足：对于每一个属于集合 Ω_0 的 θ，取值 $\pi(\theta|\delta)=0$，而对于每个属于集合 Ω_1 的 θ，$\pi(\theta|\delta)=1$. 如果检验的功效函数确实有这些值，那么不论 θ 的真实取值是什么，检验都能以概率 1 作出正确的判断. 然而在实际问题

中，检验过程中几乎很少有这种理想化的功效函数.

由于实际上在每一个检验问题中存在错误的可能性，因此应当考虑到可能犯什么样的错误. 对于 $\theta \in \Omega_0$，拒绝 H_0 是一个不正确的决定，这就是传统意义上所说的第一类错误. 所以当 $\theta \in \Omega_0$ 时，$\pi(\theta|\delta)$ 就是统计人员犯第一类错误的概率. 对于 $\theta \in \Omega_1$，接受原假设 H_0 是另一种不正确的决定，即所谓的第二类错误. 所以如果 $\theta \in \Omega_1$，$1-\pi(\theta|\delta)$ 就是犯第二类错误的概率. 当然 θ 要么属于集合 Ω_0，要么属于集合 Ω_1，不能同时属于两个集合. 因此，只有可能犯一种错误，但我们不知道是哪一种错误.

如果要在几个检验中作选择，我们会选择犯错误的概率小的检验 δ. 也就是，希望当值 $\theta \in \Omega_0$ 时，功效函数 $\pi(\theta|\delta)$ 较小，当 $\theta \in \Omega_1$ 时，$\pi(\theta|\delta)$ 将较大. 通常这两个目标是互相对立的. 即，当 $\theta \in \Omega_0$ 时，我们选择一个让 $\pi(\theta|\delta)$ 小的 δ，通常也会发现当 $\theta \in \Omega_1$ 时，$\pi(\theta|\delta)$ 是小的. 例如，如果不管是什么观测数据，检验过程 δ_0 总是接受 H_0，那么对于所有的 $\theta \in \Omega_0$，将有 $\pi(\theta|\delta_0)=0$. 然而，对于所有的 $\theta \in \Omega_1$，$\pi(\theta|\delta_0)=0$ 也会成立. 类似地，如果检验 δ_1 总是接受 H_1，那么对于所有的 $\theta \in \Omega_1$，将有 $\pi(\theta|\delta_1)=1$. 但与此同时 $\theta \in \Omega_0$，也将导致 $\pi(\theta|\delta_1)=1$. 因此，有必要寻求在 Ω_0 中低功效和 Ω_1 中高功效的两个目标之间的一个适当的平衡点.

452

在这两个目标之间达到平衡的最普通的方法就是选择一个 0 到 1 之间的数 α_0，使它满足

$$\pi(\theta|\delta) \leqslant \alpha_0, \text{对于所有的 } \theta \in \Omega_0. \tag{8.1.4}$$

那么，在所有满足(8.1.4)的检验中，统计人员寻找一个检验，它的功效函数和 $\theta \in \Omega_1$ 时得到的一样大. 将在8.2节和8.3节中讨论这种方法. 一个满足(8.1.4)的检验叫做 α_0 水平检验，称 α_0 为检验的显著性水平. 另外，δ 检验的检验水平 $\alpha(\delta)$ 定义如下：

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta|\delta). \tag{8.1.5}$$

于是当且仅当 $\alpha(\delta) \leqslant \alpha_0$，$\delta$ 是一个 α_0 水平检验. 如果原假设是简单的，即 $H_0: \theta = \theta_0$，那么 δ 的检验水平为 $\alpha(\delta) = \pi(\theta_0|\delta)$.

例 8.1.3 关于均匀分布的假设检验 设随机样本 $X_1, \cdots, X_n$ 来自于区间 $[0,\theta]$ 上的均匀分布，且 θ 的值未知($\theta>0$)，再假定要检验以下的假设：

$$\begin{aligned} &H_0: 3 \leqslant \theta \leqslant 4, \\ &H_1: \theta < 3 \text{ 或 } \theta > 4. \end{aligned} \tag{8.1.6}$$

从例6.5.15可知 θ 的M.L.E是 $Y_n = \max(X_1, \cdots, X_n)$. 虽然 Y_n 必然比 θ 小，但当样本容量 n 相当大时，Y_n 接近于 θ 还是有很大的发生概率的. 为了说明起见，假定当 $2.9 < Y_n < 4$ 时检验 δ 接受 H_0，当 Y_n 不在这一区间时拒绝 H_0. 这样一来，检验 δ 的临界区域包括所有 $Y_n \leqslant 2.9$ 或 $Y_n \geqslant 4$ 的 $X_1, \cdots, X_n$ 的值.

δ 的功效函数满足关系式

$$\pi(\theta|\delta) = \Pr(Y_n \leqslant 2.9|\theta) + \Pr(Y_n \geqslant 4|\theta).$$

如果 $\theta\leqslant 2.9$，那么有 $\Pr(Y_n\leqslant 2.9|\theta)=1$ 和 $\Pr(Y_n\geqslant 4|\theta)=0$. 因此，$\pi(\theta|\delta)=1$. 如果 $2.9<\theta\leqslant 4$，那么有 $\Pr(Y_n\leqslant 2.9|\theta)=(2.9/\theta)^n$ 且 $\Pr(Y_n\geqslant 4|\theta)=0$. 在这种情况下，$\pi(\theta|\delta)=(2.9/\theta)^n$. 最后，如果 $\theta>4$，那么有 $\Pr(Y_n\leqslant 2.9|\theta)=(2.9/\theta)^n$ 且 $\Pr(Y_n\geqslant 4|\theta)=1-(4/\theta)^n$. 在这种情况下，$\pi(\theta|\delta)=(2.9/\theta)^n+1-(4/\theta)^n$. 功效函数$\pi(\theta|\delta)$的草图由图 8-1 表示.

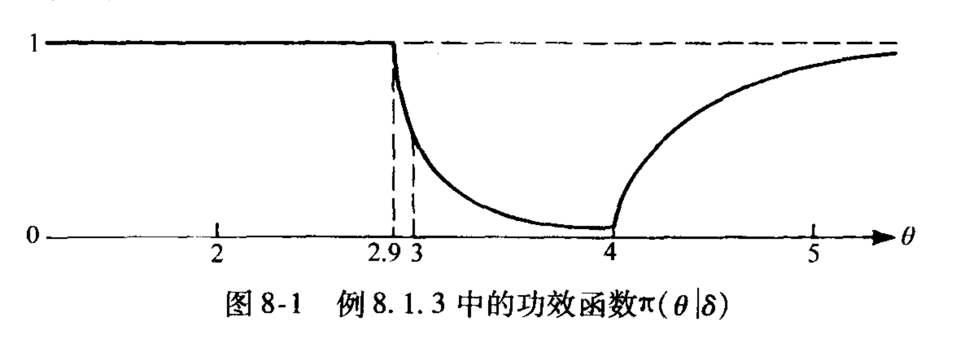


图 8-1　例 8.1.3 中的功效函数$\pi(\theta|\delta)$ ◀

根据等式(8.1.5)可知，δ 的检验水平为$\sup_{3\leqslant\theta\leqslant 4}\pi(\theta|\delta)$. 由图 8-1 和上面所做的运算可以得到 $\alpha(\delta)=\pi(3|\delta)=(29/30)^n$. 特别地，如果样本容量 $n=68$，那么 δ 的检验水平为 $(29/30)^{68}=0.0997$. 所以，δ 是一个对于每一个显著性水平 $\alpha_0\geqslant 0.0997$ 的 α_0 水平检验.

453

8.1.5 使检验有指定的显著性水平

假定我们希望检验假设

$$H_0: \theta\in\Omega_0,$$
$$H_1: \theta\in\Omega_1.$$

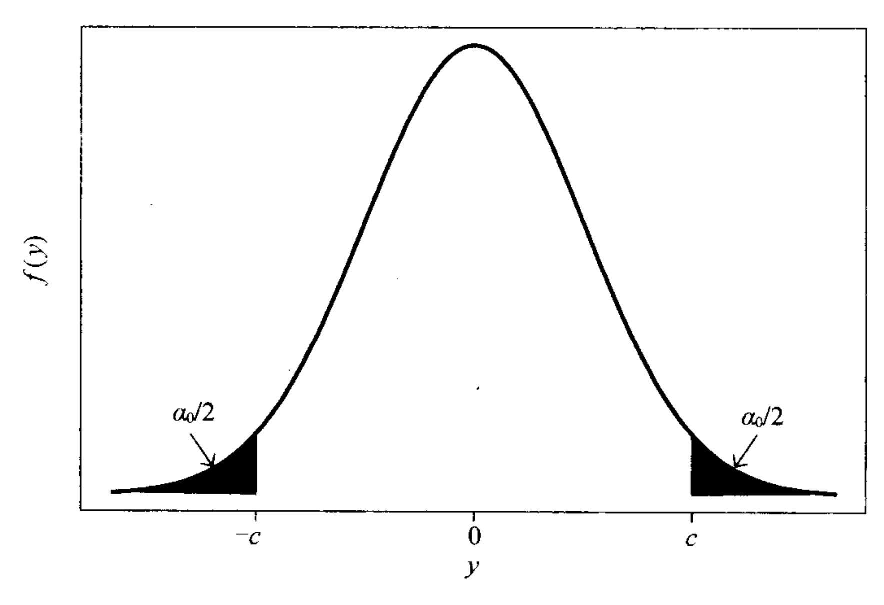
设 T 为一个检验统计量，假定对于常数 c，如果 $T\geqslant c$，那么我们的检验就拒绝原假设. 再假定希望此检验的显著性水平为 α_0，检验的功效函数为$\pi(\theta|\delta)=\Pr(T\geqslant c|\theta)$，我们想要

$$\sup_{\theta\in\Omega_0}\Pr(T\geqslant c|\theta)\leqslant\alpha_0. \tag{8.1.7}$$

很清楚地，功效函数和(8.1.7)的左边是 c 的非增函数. 因此，要想使(8.1.7)成立，c 的值就要大，而不能小. 如果想让 $\theta\in\Omega_1$ 时，功效函数尽可能的大，就应该让 c 在满足(8.1.7)的情况下尽可能地小. 如果 T 有一个连续分布，那么很容易就能找到一个合适的 c.

例 8.1.4　检验关于有已知方差的正态分布的均值的假设　在例 8.1.2 中，当 $|\overline{X}_n-\mu_0|\geqslant c$ 时，我们的检验拒绝 $H_0:\mu=\mu_0$. 由于原假设很简单，所以(8.1.7)的左边将简化为 $|\overline{X}_n-\mu_0|\geqslant c$ 的概率(假设 $\mu=\mu_0$). 由于当 $\mu=\mu_0$ 时，$Y=\overline{X}_n-\mu_0$ 有均值为 0、方差为 σ^2/n 的正态分布，所以对于每一个 α_0 都可以找到一个使检验水平恰好为 α_0 的 c 值. 图 8-2 显示了 Y 的概率密度函数和在此概率密度函数下阴影部分所

表示的检验水平. 由于正态分布的概率密度函数是关于均值(这里是0)对称的，所以两个阴影的面积是一样的，即为 $\alpha_0/2$. 这就意味着 c 必须是 Y 的分布中的 $1-\alpha_0/2$ 分位数. 该分位数就是 $c=\Phi^{-1}(1-\alpha_0/2)\sigma n^{-1/2}$.



454

图8-2 例8.1.4中 $Y=\overline{X}_n-\mu_0$ 当 $\mu=\mu_0$ 时的概率密度函数. 阴影部分为 $|Y|\geq c$ 的概率

当检验有关正态分布的均值的假设时，通常的做法是用统计量

$$Z=n^{1/2}\frac{\overline{X}_n-\mu_0}{\sigma}. \tag{8.1.8}$$

来改写这个检验. 如果 $|Z|\geq\Phi^{-1}(1-\alpha_0/2)$，那么检验拒绝 H_0. ◀

例8.1.5 检验关于伯努利参数的假设 假定 $X_1,\cdots,X_n$ 是独立同分布的且服从参数为 p 的伯努利分布. 假定我们希望检验假设

$$\begin{aligned}&H_0: p\leq p_0,\\&H_1: p>p_0.\end{aligned} \tag{8.1.9}$$

令 $Y=\sum_{i=1}^{n}X_i$，它服从参数为 n 和 p 的二项分布. 希望 p 越大时 Y 就越大. 所以，假定对于某一常数 c，当 $Y\geq c$ 时，选择拒绝 H_0. 再假定检验水平在不超过 α_0 的前提下尽可能地接近 α_0. 很容易验证 $\Pr(Y\geq c|p)$ 是 p 的增函数，因此检验水平就是 $\Pr(Y\geq c|p=p_0)$. 所以，c 是满足使 $\Pr(Y\geq c|p=p_0)\leq\alpha_0$ 达到最小的一个数. 例如，当 $n=10, p_0=0.3, \alpha_0=0.1$ 时，可以利用书后的二项概率表来确定 c. 我们可以计算得到 $\sum_{y=6}^{10}\Pr(Y=y|p=0.3)=0.0473$ 和 $\sum_{y=5}^{10}\Pr(Y=y|p=0.3)=0.1503$. 为了保持至多0.1的检验水平，必须选择 $c>5$. 每一个在区间 $(5,6]$ 上的 c 值产生的是同一个检验，因为 Y 只取整数值. ◀

不管何时，当选择一个检验过程时，需要检验它的功效函数. 当你做了一个好

选择的时候，则对于 $\theta\in\Omega_1$ 而言功效函数应当比对于 $\theta\in\Omega_0$ 要大些．再者，当 θ 不在 Ω_0 范围内时，功效函数应当增加．例如，图 8-3 显示了本节中的两个例子中的功效函数的图形.

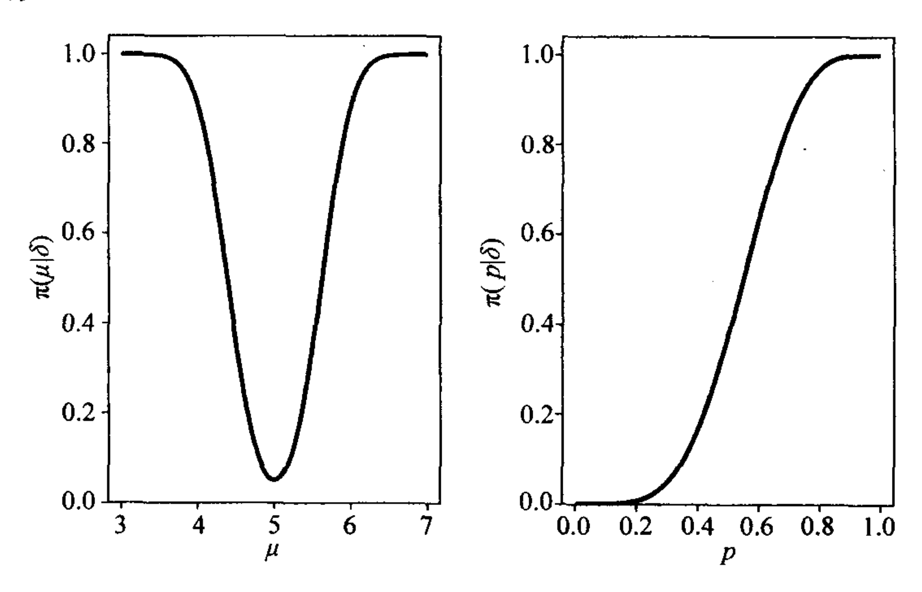


图 8-3　两个检验的功效函数．左边的图是例 8.1.4 检验的功效函数，$n=10$，$\mu_0=5$，$\sigma=1$ 和 $\alpha_0=0.05$．右边的图是例 8.1.5 检验的功效函数，$n=10$，$p_0=0.3$ 和 $\alpha_0=0.1$

8.1.6 p-值

假设检验的结果可能会出现数据未得到充分使用的情况．例如，在例 8.1.4 中，假如进行 $\alpha=0.05$ 的检验．那么如果在等式(8.1.8)的统计量 Z 至少为 1.96 时，应该拒绝 H_0．即，无论是 $Z=1.97$ 或者 $Z=6.97$，我们将得到同样的结果，也就是在 0.05 的水平下，拒绝 H_0．该检验结果不与其他任何决定有关．进一步，如果使用 $\alpha_0=0.01$的检验，那么在 $Z=1.97$ 下不能拒绝 H_0，在 $Z=6.97$ 下，将拒绝 H_0.

根据这些原因，一个实验者在实际过程中不会事先选择 α_0 的值，而后再简单地报告在 α_0 的水平下是否接受或拒绝 H_0．在许多应用领域，除了给出像 Z 这样的适当的检验统计量的观察值以外，人们已经习惯于给出在 α_0 检测水平下将拒绝 H_0 的所有的 α_0 值．这样一来，如果在例 8.1.4 中 Z 的值为 1.97，那么对于每个满足 $1.97\geqslant\Phi^{-1}(1-\alpha_0/2)$ 的显著性水平 α_0，H_0 将被拒绝．使用在书后给出的正态分布表，该不等式转化为 $\alpha_0\geqslant 0.0488$．称 0.0488 为相应观察数据和假设检验的 p-值．通常，p-值是我们根据观察数据在 α_0 水平下拒绝原假设的最小的 α_0 水平. 455

在实验的分析中，实验者通常会报告说，Z 的观察值为 1.97，而相应的 p-值为 0.0488，也就是说在 0.0488 的显著性水平下，观察值 Z 刚好显著．当实验者以这样方法报告实验结果时，有利的一点是他不必事先为将要进行的检验选择任何一个 α_0 的显著性水平．另外，当该实验报告的一个读者看到 Z 的取值在 0.0488 的显著性水平下正好显著，他立刻就知道对于比较大的 α_0 值，应该拒绝 H_0，对于较小的 α_0 值

不应拒绝 H_0.

对于这种"当 $T \geq c$ 时，拒绝原假设"的形式的检验，有一种直接的方法去计算 p-值. 对于每个 t，令 δ_t 是如果 $T \geq t$ 则拒绝 H_0 的检验. 那么当 $T = t$ 时 P 的值就是 δ_t 检验的水平（见习题18），也就是说 p-值等于

$$\sup_{\theta \in \Omega_0} \pi(\theta | \delta_t) = \sup_{\theta \in \Omega_0} \Pr(T \geq t | \theta). \tag{8.1.10}$$

通常，在 Ω_0 到 Ω_1 的边界上的某一个 θ_0，会使 $\pi(\theta|\delta_t)$ 最大. 因为 p-值是在 T 分布的上尾部作为概率来计算的，有时就称它为*尾域*.

456

例 8.1.6　检验关于伯努利参数的假设　对于在例 8.1.5 中对于假设(8.1.9)的检验，我们使用若 $Y \geq c$，则拒绝 H_0 的检验. 当 $Y = y$ 出现时，p-值将等于 $\sup_{p \leq p_0} \Pr(Y \geq y | p)$. 在本例中，很容易看到 $\Pr(Y \geq y | p)$ 作为 p 的函数是递增的，因此，p-值是 $\Pr(Y \geq y | p = p_0)$. 例如，令 $p_0 = 0.3$，$n = 10$. 如果 $Y = 6$，那么 $\Pr(Y \geq 6 | p = 0.3) = 0.0473$，这正是我们在例 8.1.5 所得到的. ◀

当检验没有这种"如果 $T \geq c$，拒绝 H_0"的形式，p-值的定义比较复杂. 在本书中，我们仅计算这种形式的检验的 p-值.

8.1.7　检验和置信集合的等价性

接下来，将说明置信区间(见 7.5 节)是如何作为另外一种方法来报告假设检验的结果的. 特别地，将说明可以将一个系数 γ 的置信集(置信区间的推广)看作在显著性水平 $1-\gamma$ 下被接受的原假设的集合.

准确地说，假定一个随机样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是取自一个含有未知参数 θ 的分布. 已知该参数 θ 属于参数空间 Ω，但 θ 可能是一个实数或者一个向量. 对于任意给定的点 $\theta_0 \in \Omega$，考虑以下的检验假设：

$$\begin{aligned} &H_0: \theta = \theta_0, \\ &H_1: \theta \neq \theta_0. \end{aligned} \tag{8.1.11}$$

假如对于每个 $\theta_0 \in \Omega$ 和每个 $\gamma(0 < \gamma < 1)$，能构造置信水平为 $1-\gamma$ 的基于假设(8.1.11)的检验 δ_{θ_0}. 即能构造一个检验 δ_{θ_0}，符合如下条件：

$$\Pr[\text{当使用检验 } \delta_{\theta_0} \text{ 时拒绝 } H_0 | \theta = \theta_0] \leq 1 - \gamma. \tag{8.1.12}$$

对于在随机样本中可能观察到的数值 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 的每一个集合，令 $\omega(\boldsymbol{x})$ 表示所有那些点 $\theta_0 \in \Omega$ 的集合，与这些 θ_0 对应的检验 δ_{θ_0} 在观察到数据 $\boldsymbol{X} = \boldsymbol{x}$ 时接受假设 H_0. 集合 $\omega(\boldsymbol{x})$ 是参数空间 Ω 的子集，并且可看作是对 θ 的一个集合估计. 事实上，如果将 $\omega(\boldsymbol{X})$ 看作一个在观察数据之前的随机集合时，那么将证明下式：

$$\Pr[\theta_0 \in \omega(\boldsymbol{X}) | \theta = \theta_0] \geq \gamma. \tag{8.1.13}$$

基于对 $\omega(\boldsymbol{x})$ 的定义，一个给定的点 θ_0 落在子集 $\omega(\boldsymbol{x})$ 中当且仅当观察数据为 $\boldsymbol{X} = \boldsymbol{x}$ 时检验 δ_{θ_0} 接受原假设 $H_0: \theta = \theta_0$. 因此，由等式(8.1.12)有

$$\Pr[\theta_0 \in \omega(\boldsymbol{X}) | \theta = \theta_0] = \Pr[\text{当使用检验 } \delta_{\theta_0} \text{ 时接受 } H_0 | \theta = \theta_0] \geq \gamma. \tag{8.1.14}$$

若对于每个 $\theta_0 \in \Omega$，一个随机集合 $\omega(\boldsymbol{X})$ 都满足(8.1.13)，那么就称 $\omega(x)$ 是对 θ 的系数 γ 置信集．若对于所有 θ_0，不等式(8.1.13)都成为等式，那就称该置信集合是正合的（或称精确的）．刚才定义的置信集合，是对在 7.5 节中介绍的置信区间概念的一般化．我们所证明的就是假设(8.1.11)的一族置信水平 $1-\gamma$ 检验可以用来构造 θ 的系数 γ 置信集合．

相反地，假定 $\omega(\boldsymbol{x})$ 是对于 θ 的有置信系数 γ 的置信集合，那么对于每一个 $\theta_0 \in \Omega$，定义检验步骤 $\delta^*_{\theta_0}$ 如下：当且仅当 $\theta_0 \in \omega(\boldsymbol{x})$，接受检验 H_0．我们很容易发现对于每个 θ_0 来说 $\delta^*_{\theta_0}$ 是在水平为 $1-\gamma$ 的检验．因此，我们已经证明了对于 θ 构造一个有置信系数 γ 的置信集合等价于对于每个 $\theta_0 \in \Omega$ 构造假设(8.1.11)的一族水平为 $1-\gamma$ 的检验．

例 8.1.7　正态分布均值的置信区间　考虑在例 8.1.14 中出现的对于假设(8.1.2)的检验．令 $\alpha_0 = 1-\gamma$，若 $|\overline{X}_n - \mu_0| \geqslant \Phi^{-1}(1-\alpha_0/2)\sigma n^{-1/2}$，$\alpha_0$ 水平的检验 δ_{μ_0} 将拒绝 H_0．若出现 $\overline{X}_n = \overline{x}_n$，接受 H_0 的 μ_0 集合如下

$$|\overline{x}_n - \mu_0| < \Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\sigma n^{-1/2}.$$

该不等式可方便地变为

$$\overline{x}_n - \Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\sigma n^{-1/2} < \mu_0 < \overline{x}_n + \Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\sigma n^{-1/2}.$$

系数 γ 置信区间成为 $(\overline{x}_n - \Phi^{-1}(1-\alpha_0/2)\sigma n^{-1/2}, \overline{x}_n + \Phi^{-1}(1-\alpha_0/2)\sigma n^{-1/2})$．对于所有的 θ_0，容易证实

$$\Pr\left(\theta_0 \in \left(\overline{X}_n - \Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\sigma n^{-1/2}, \overline{X}_n + \Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\sigma n^{-1/2}\right) \middle| \theta = \theta_0\right) = \gamma.$$

该置信区间是正合的．◀

例 8.1.8　由置信区间构造一个检验　在 7.5 节中，我们学习了当面对均值与方差均未知的正态分布时，如何构造置信区间．令 $X_1, \cdots, X_n$ 为取自均值与方差均未知的正态分布的随机样本．在 7.5 节中使用了统计量

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \sigma' = \left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right)^{1/2}. \tag{8.1.15}$$

系数 γ 置信区间为以下区间的观察值

$$\left(\overline{X}_n - \frac{c}{n^{1/2}}\sigma', \overline{X}_n + \frac{c}{n^{1/2}}\sigma'\right), \tag{8.1.16}$$

其中 c 为服从自由度为 $n-1$ 的 t 分布的 $(1+\gamma)/2$ 分位数．对于每一个 μ_0，可以用该区间得到在水平 $\alpha_0 = 1-\gamma$ 下的假设检验

458

$$H_0: \mu = \mu_0,$$
$$H_1: \mu \neq \mu_0.$$

若 μ_0 不在区间(8.1.16)中，此检验将拒绝 H_0. 稍许做一点代数运算就能发现当且仅当

$$\left| n^{1/2}\frac{\overline{X}_n - \mu_0}{\sigma'} \right| \geq c$$

时，μ_0 不在区间(8.1.16)内. 该检验等价于我们将在8.3节中更详细学习的 t 检验.

单侧置信区间

存在着和单侧假设检验相对应的单侧置信区间. 设 θ 是一个参数，并且令 $g(\theta)$ 为 θ 的一维函数. 例如，如果 $\theta = (\mu, \sigma^2)$ 是一族正态分布的参数，则 $g(\theta) = \mu$ 就是一个一维函数. 如果对于每一个 g_0，有一个以下假设的水平 α_0 检验：

$$H_0: g(\theta) \leq g_0,$$
$$H_1: g(\theta) > g_0. \tag{8.1.17}$$

（或者有相反的不等式的情形）那么可以用这些检验形成一个单侧置信区间. 该区间将包含根据观察数据将接受原假设 H_0 的所有的 g_0，而且它是系数 $1-\alpha_0$ 的置信区间.

例8.1.9　伯努利参数的单侧置信区间　在例8.1.5中，已经知道如何在 α_0 水平下构造单侧假设检验. 设 $Y = \sum_{i=1}^{n} X_i$. 如果 $Y \geq c(p_0)$，其中 $c(p_0)$ 是使 $P(Y \geq c \mid p = p_0) \leq \alpha_0$ 的最小的 $c(p_0)$，那么就拒绝原假设 H_0. 在观测到数据 $\boldsymbol{X}$ 后，就可以对于每个 p_0 验证是否拒绝 H_0，即，对于每个 p_0 都验证是否有 $Y \geq c(p_0)$. 所有满足 $Y < c(p_0)$（即接受 H_0）的 p_0 将形成一个区间 $\omega(\boldsymbol{X})$，该区间对于所有的 p_0 都满足 $P(p_0 \in \omega(\boldsymbol{X}) \mid p = p_0) \geq 1 - \alpha_0$. 例如，假定 $n = 10, \alpha_0 = 0.1$，并且观察到 $Y = 6$，为了在0.1的检验水平下接受 $H_0: p \leq p_0$，必须有一个不包含6的临界区域，而该种情况当且仅当 $P(Y \geq 6 \mid p = p_0) > 0.1$ 时才会发生. 通过尝试 p_0 的不同的值，发现对于所有的 $p_0 > 0.3542$ 的值此不等式 $P(Y \geq 6 \mid p = p_0) > 0.1$ 都成立. 因此，如果观察到 $Y = 6$，那么系数0.9的置信区间是（0.3542，1）. 注意，0.3不在此区间内，所以在0.1的检验水平下拒绝 $H_0: p \leq 0.3$，就像在例8.1.5时所做的那样，对于其他观测值 $Y = y$，置信区间都将是 $(q(y), 1)$ 的形式. 这里 $q(y)$ 的计算方法将在习题17中做简要介绍. 当 $n = 10$，$\alpha_0 = 0.1$ 时 $q(y)$ 的值是：

y	0	1	2	3	4	5	6	7	8	9	10
$q(y)$	0	0.0104	0.0545	0.1158	0.1875	0.2673	0.3542	0.4482	0.5503	0.6631	0.7943

这些置信区间都不是正合的.

一个以 $1-\alpha_0$ 为系数的单侧置信区间 $\omega(\boldsymbol{x})$ 可以用来构造一个假设(8.1.17)的检验 δ. 如果 g_0 不在区间 $\omega(\boldsymbol{x})$ 内，就拒绝 H_0. 虽然这样的检验 δ 满足对于使 $g(\theta) = g_0$

成立的所有 θ，都有$\pi(\theta|\delta)\leqslant\alpha_0$，而这也不能说明 δ 有水平 α_0，即，在每个具体的案例中，仍然有必要检查是否对于所有满足 $g(\theta)\leqslant g_0$ 的 θ 有$\pi(\theta|\delta)\leqslant\alpha_0$.

459

例 8.1.10 方差未知的正态均值的单侧检验和置信区间 设 $X_1,\cdots,X_n$ 是一个来自于方差 σ^2 和均值 μ 都未知的正态分布的随机样本，这里 $\theta=(\mu,\sigma^2)$. 令 $g(\theta)=\mu$. 在 7.4 节中，已经知道随机变量

$$U=\frac{n^{1/2}(\overline{X}_n-\mu)}{\sigma'} \tag{8.1.18}$$

服从自由度为 $n-1$ 的 t 分布，这里 $\overline{X}_n$ 和 σ'在(8.1.15)中已经有定义. 令 T_{n-1}^{-1}是该 t 分布的分位函数. 它满足 $\Pr(U\geqslant T_{n-1}^{-1}(1-\alpha_0)|\mu,\sigma^2)=\alpha_0$. 该等式等价于:

$$\Pr(\mu\geqslant\overline{X}_n-\sigma' n^{-1/2}T_{n-1}^{-1}(1-\alpha_0)|\mu,\sigma^2)=1-\alpha_0. \tag{8.1.19}$$

(8.1.19)说明了 μ 的以 $1-\alpha_0$ 为参数的单侧置信区间是如下区间的观察值:

$$(\overline{X}_n-\sigma' n^{-1/2}T_{n-1}^{-1}(1-\alpha_0),\infty). \tag{8.1.20}$$

现在假定用该区间去检验假设. 如果 μ_0 不在(8.1.20)区间内，应该拒绝原假设 $\mu=\mu_0$. 很容易看到，当且仅当 $\overline{X}_n\geqslant\mu_0+\sigma' n^{-1/2}T_{n-1}^{-1}(1-\alpha_0)$时，$\mu_0$ 不在区间(8.1.20). 这样的检验似乎对于检验以下的假设是有意义的.

$$H_0:\mu\leqslant\mu_0,$$
$$H_1:\mu>\mu_0.$$

以上构造的检验是我们将在 8.3 节中研究的另一种 t 检验. 特别地，将在 8.3 节中证明 t 检验是一个 α_0 水平的检验. 在习题 19 中你会遇到对应于检验相反假设的单侧置信区间. ◀

8.1.8 小结

假设检验是研究确定参数 θ 是否落在参数空间的一个特定的子集合 Ω_0 中或它的补集 Ω_1 中的问题，我们把 $\theta\in\Omega_0$ 的情况称为原假设并表示为 H_0，而另外一种情况称为备择假设并表示为 $H_1:\theta\in\Omega_1$. 设 S 是可能观测到的所有可能的数据值或向量的集合，观测值可分为两个互不相交的子集，当观测值 $X\in C(C\subset S)$时拒绝原假设 H_0，当观测值 $X\notin C(X\in C$ 的补集)时接受原假设 H_0，则称这样的子集 C 为检验的临界域. 而检验的功效函数 δ 就是$\pi(\theta|\delta)=\Pr(X\in C|\theta)$，$\delta$ 的水平是 $\sup_{\theta\in\Omega_0}\pi(\theta|\delta)$，当此值最多为 α_0 时，就把检验称为 α_0 水平的，如果 Ω_0 是一个参数观测值的集合，则称原假设 H_0 为简单假设. 反之 H_0 为复合假设. 类似地，如果 Ω_1 是一个单点集时，称 H_1 为简单假设，反之则为复合假设，当 H_0 为真时而拒绝了 H_0，称为第一类错误，而当 H_1 为真，接受 H_0 时，称为第二类错误.

假设检验通常是建立一个统计量 T 来进行检验，当统计量 T 落在一定区间内或一定区间外就拒绝原假设. 该区间的划分使检验有了期望的显著性水平. p-值能使我们

460

的结果提供更多的信息,只要是采用检验形式“如果某统计量 $T\geq c$ 时拒绝 H_0”,就能容易地计算出 p-值来. 当 $T=t$ 时,p-值就等于 $\sup_{\theta\in\Omega_0}\Pr(T\geq t|\theta)$. 我们还说明了为什么置信集合是反映假设检验所得结果的一种方式. 当以显著性水平 α_0 接受原假设 $H_0: \theta=\theta_0$时,以置信系数 $1-\alpha_0$ 给出的置信区间就是所有 $\theta_0\in\Omega$ 的集合. 当我们检验关于一维参数或参数的一维函数的假设时,这些置信集合是区间.

习题

1. X 服从参数为 β 的指数分布,假定要检验假设

$$H_0: \beta\geq 1,$$
$$H_1: \beta<1.$$

考虑当 $X\geq 1$ 时拒绝 H_0 的检验步骤 δ.

a. 确定检验的功效函数;

b. 计算检验的水平.

2. 若 $X_1,\cdots,X_n$ 是$[0,\theta]$上均匀分布的一个随机样本，检验假设

$$H_0: \theta\geq 2,$$
$$H_1: \theta<2.$$

令 $Y_n=\max(X_1,\cdots,X_n)$，临界域包含使 $Y_n\leq 1.5$ 的所有结果. 考虑相应的检验过程.

a. 写出检验的功效函数;

b. 求检验的水平.

3. 假定在一大批物品中残次品所占的比例是未知的，要检验假设

$$H_0: p=0.2,$$
$$H_1: p\neq 0.2.$$

并设从这些物品中随机取出 20 个作为随机样本，令 Y 表示该样本中残次品的件数，临界域包含了所有使 $Y\leq 1$ 或 $Y\geq 7$ 的结果，考虑相应的检验步骤 δ.

a. 计算出当 $p=0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1$ 时的功效函数$\pi(p|\delta)$的值并画出功效函数的草图;

b. 求检验的水平.

4. 若 $X_1,\cdots,X_n$ 是服从均值 μ 为未知、方差为 1 的正态分布的一个随机样本，μ_0 为一个确定常数，检验假设

$$H_0: \mu=\mu_0,$$
$$H_1: \mu\neq\mu_0.$$

并设样本容量 n 为 25，考虑当 $|\overline{X}_n-\mu_0|<c$ 则接受 H_0 的检验方法. 计算使检验的显著性水平为 0.05 的 c 值.

5. 假设 $X_1,\cdots,X_n$ 是服从均值与方差都未知的正态分布的一个随机样本，区分以下假设是简单假设还是复合假设.

a. $H_0: \mu=0$ 且 $\sigma=1$.

b. $H_0: \mu>3$ 且 $\sigma<1$.

c. $H_0: \mu=-2$ 且 $\sigma^2<5$.

d. $H_0: \mu = 0$.

6. 若 X 服从 $\left[\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]$ 上的均匀分布，检验假设

$$H_0: \theta \leqslant 3,$$
$$H_1: \theta > 4.$$

构造检验过程 δ 使功效函数有以下值：当 $\theta \leqslant 3$ 时，$\pi(\theta|\delta) = 0$；当 $\theta \geqslant 4$ 时，$\pi(\theta|\delta) = 1$.

7. 回顾以前的例子 8.1.3，考虑一个新的检验 δ^*，使得当 $2.9 \leqslant Y_n \leqslant 4.5$ 时接受 H_0，设 δ 是例 8.1.3 给出的检验步骤

a. 证明：当 $\theta \leqslant 4$ 时，$\pi(\theta|\delta^*) = \pi(\theta|\delta)$.

b. 证明：当 $\theta > 4$ 时，$\pi(\theta|\delta^*) < \pi(\theta|\delta)$.

c. 在两个检验中哪一个能较好地检验假设(8.1.6)？

461

8. 假设 $X_1, \cdots, X_n$ 独立同分布于一个平均值是 μ、方差是 1 的正态分布，检验假设：

$$H_0: \mu \leqslant \mu_0,$$
$$H_1: \mu > \mu_0.$$

考虑检验如果 $Z \geqslant c$，则拒绝 H_0，此处 Z 由(8.1.8)式确定.

a. 说明 $\Pr(Z \geqslant c \mid \mu)$ 是 μ 的递增函数；

b. 确定常数 c，使检验的显著性水平为 α_0.

9. 假设 $X_1, \cdots, X_n$ 独立同分布于一个平均值是 μ、方差是 1 的正态分布，检验假设：

$$H_0: \mu \geqslant \mu_0,$$
$$H_1: \mu < \mu_0.$$

找一个检验统计量 T，使得对于每一个 c，当 $T \geqslant c$ 时拒绝 H_0 的检验 δ_c 都有随 μ 递减的功效函数 $\pi(\mu|\delta_c)$.

10. 在习题 8 中，假设 $Z = z$，试确定 p-值的公式.

11. 设 $X_1, \cdots, X_9$ 独立同分布，有参数为 p 的伯努利分布，检验假设：

$$H_0: p = 0.4,$$
$$H_1: p \neq 0.4.$$

令 $Y = \sum_{i=1}^{9} X_i$.

a. 试确定 c_1 和 c_2，使

$$\Pr(Y \leqslant c_1 \mid p = 0.4) + \Pr(Y \geqslant c_2 \mid p = 0.4)$$

不大于但尽可能接近于 0.1.

b. 设 δ 是如果 $Y \leqslant c_1$ 或者 $Y \geqslant c_2$，则拒绝 H_0 的检验过程. 求 δ_c 的检验水平.

c. 画出检验函数 δ_c 的图像.

12. 设 X 是一个中心是 θ 的柯西分布的观察值，X 的概率密度函数是

$$f(x|\theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \quad 对于 -\infty < x < \infty.$$

检验假设：

$$H_0: \theta \leqslant \theta_0,$$
$$H_1: \theta > \theta_0.$$

检验过程 δ_c 为当 $X\geqslant c$，则拒绝 H_0.

a. 证明：$\pi(\theta\mid\delta_c)$ 是 θ 的递增函数；

b. 确定常数 c 使检验水平为 0.05；

c. 假设 $X=x$，试确定 p-值的公式.

13. 设 X 是服从平均值为 θ 的泊松分布，检验假设：

$$H_0: \theta\leqslant 1.0,$$
$$H_1: \theta > 1.0.$$

将检验 δ_c 规定为：如果 $X\geqslant c$，则拒绝 H_0. 试确定常数 c，使 δ_c 的检验水平不大于 0.1 且尽可能接近于 0.1.

14. 设 $X_1,\cdots,X_n$ 独立同分布于参数为 θ 的指数分布，检验假设：

$$H_0: \theta\geqslant\theta_0,$$
$$H_1: \theta<\theta_0.$$

令 $X=\sum_{i=1}^n X_i$，δ_c 是如果 $X\geqslant c$，则拒绝 H_0 的检验.

a. 证明：$\pi(\theta\mid\delta_c)$ 是 θ 的递减函数；

b. 确定常数 c，使检验 δ_c 的显著性水平为 α_0；

c. 令 $\theta_0=2$，$n=1$ 和 $\alpha_0=0.1$，找到 δ_c 的精确形式并画出它的功效函数的草图.

15. 设 X 服从区间 $[0,\theta]$ 上的均匀分布，要检验以下假设：

$$H_0: \theta\leqslant 1,$$
$$H_1: \theta>1.$$

考虑有下述形式的检验过程："当 $X\geqslant c$，则拒绝 H_0"，对于 X 的每个可能的值 x，求观察到 $X=x$ 时的 p-值.

16. 考虑 7.5 节习题 5 中的置信区间，确定与该区间等价的假设检验集合，即：对于每一个 $c>0$，找一个检验 δ_c，使检验原假设 $H_{0,c}$ 为 $\sigma^2=c$，而备择假设导致当且仅当 c 不在区间内时，δ_c 拒绝 $H_{0,c}$，通过在依赖 c 的某个非随机区间的内部或外部的统计量 $T=r(\boldsymbol{X})$ 来写出该检验.

17. 设 $X_1,\cdots,X_n$ 独立同分布，且有参数为 p 的伯努利分布. 令 $Y=\sum_{i=1}^n X_i$，要确定 p 的形式为 $(q(y),1)$ 系数 γ 的置信区间，证明：如果 $Y=y$，则 $q(y)$ 应选择满足 $\Pr(Y\geqslant y\mid p=p_0)\geqslant 1-\gamma$ 的最小值 p_0.

18. 考虑在公式(8.1.10)前面部分描述的情况，证明(8.1.10)式等于置信水平 α_0 上拒绝 H_0 中最小的 α_0 值.

462

19. 回到例 8.1.10. 在 α_0 水平上检验假设：

$$H_0: \mu\geqslant\mu_0,$$
$$H_0: \mu<\mu_0.$$

$\overline{X}_n$ 很小时，拒绝 H_0 是合情合理的. 构造 μ 的一个单侧系数 $1-\alpha_0$ 的置信区间，如果 μ 不在区间内，我们就拒绝 H_0. 设法确保这样给出的检验在 $\overline{X}_n$ 小时会拒绝 H_0.

8.2 双侧的备择假设

当简单的原假设所对应的备择假设为双侧假设时，对检验过程的选择需要比单

侧情形小心一点. 本节将讨论某些有关的问题并描述最常用的一些选择. 需要选择一种比备择假设为单侧时更精确的检验过程. 本节中所讨论的一些例子都选择了最普遍的检验过程.

8.2.1 过程的一般的形式

假定变量 $X_1,\cdots,X_n$ 为来自于一个服从正态分布的随机样本，正态分布的均值 μ 未知，方差 σ^2 已知. 检验下列假设： 479

$$H_0:\mu=\mu_0,\qquad H_1:\mu\neq\mu_0. \tag{8.2.1}$$

可以证明，在任何显著性水平 $\alpha_0(0<\alpha_0<1)$ 下，不存在假设(8.2.1)的 UMP 检验. 以前所定义的检验过程都不适合用来检验假设(8.2.1)，因为这些过程都是为单侧备择假设设计的，而这里的备择假设 H_1 是双侧的. 然而，由样本均值 $\overline{X}_n$ 是 μ 的极大似然估计来看，如果 $\overline{X}_n$ 足够接近于 μ_0，则可以接受假设(8.2.1)中的 H_0，否则就拒绝 H_0. 即，使用若 $\overline{X}_n\leqslant c_1$ 或 $\overline{X}_n\geqslant c_2$，则拒绝 H_0 这样的检验过程 δ 是合理的，其中 c_1 和 c_2 为适当选择的常数.

如果检验的水平为 α_0，则 c_1 和 c_2 的值必须满足下列关系：

$$\Pr(\overline{X}_n\leqslant c_1\mid\mu=\mu_0)+\Pr(\overline{X}_n\geqslant c_2\mid\mu=\mu_0)=\alpha_0. \tag{8.2.2}$$

满足上式(8.2.2)的 c_1 和 c_2 的值有无数对，当 $\mu=\mu_0$，随机变量 $n^{1/2}(\overline{X}_n-\mu_0)/\sigma$ 服从标准正态分布. 通常，我们用 Φ 表示标准正态分布的分布函数，那么等式(8.2.2)相当于：

$$\Phi\left[\frac{n^{1/2}(c_1-\mu_0)}{\sigma}\right]+1-\Phi\left[\frac{n^{1/2}(c_2-\mu_0)}{\sigma}\right]=\alpha_0. \tag{8.2.3}$$

相应于满足 $\alpha_1+\alpha_2=\alpha_0$ 每一对正数 α_1 和 α_2，都存在一对数 c_1 和 c_2，使 $\Phi[n^{1/2}(c_1-\mu_0)/\sigma]=\alpha_1$，且 $1-\Phi[n^{1/2}(c_2-\mu_0)/\sigma]=\alpha_2$. 此处每一对 c_1 和 c_2 的值都满足等式(8.2.2). 480

8.2.2 检验过程的选择

对于一个给定容量为 n 的样本，常数 c_1 和 c_2 的选择应该使功效函数的大小和形状适合于特定问题的求解. 在某些问题中，一般希望不要拒绝原假设，除非数据强烈表明 μ 与 μ_0 有很大差异. 在这种问题中，应该用较小的 α_0. 在其他问题中，当 μ 稍大于 μ_0 时接受原假设 H_0 所犯的错误比当 μ 稍小于 μ_0 时接受 H_0 所犯的错误更为严重. 这时选择有如图 8-4 中 $\pi(\mu\mid\delta_4)$ 之类的功效函数的检验会优于选择有对称性函数的检验.

通常，对于一个已知问题选择检验方法应该综合考虑当 $\mu=\mu_0$ 时拒绝 H_0 所付出的代价以及对于任意的 μ，当 $\mu\neq\mu_0$ 时接受 H_0 所付出的代价. 另外，当选定检验时，

μ 取不同值可能性的大小也应当要考虑. 例如，如果 $\mu > \mu_0$ 的可能性大于 $\mu < \mu_0$ 时，那么选择当 $\mu > \mu_0$ 时功效函数大的，而当 $\mu < \mu_0$ 时功效函数不太大的检验比选择有相反关系的要好.

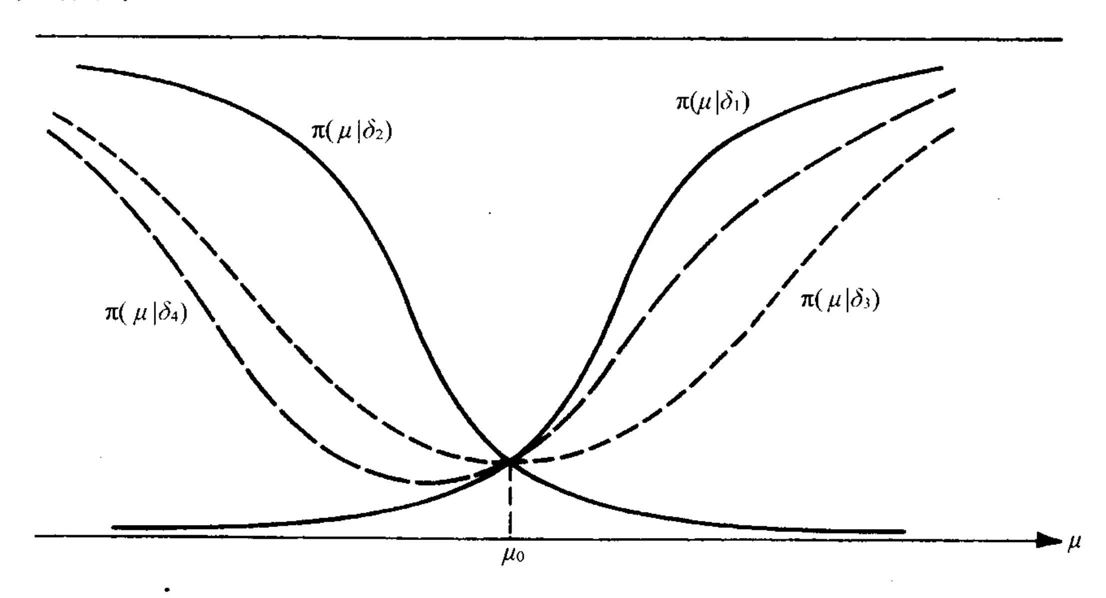


图 8-4　四个测试步骤中的功效函数

8.2.3　复合原假设

从某个角度来看，建立如(8.2.1)的假设检验是无意义的，其中原假设 H_0 对于参数 μ 指定了唯一确定的 μ_0. 在现实的问题中很难想象 μ 会那么凑巧就等于 μ_0，所以我们知道假设的 H_0 是不可能成立的. 因此. 当 H_0 给出以后，立即就遭拒绝.

这种判断仅在理论上是正确的，然而在许多问题中，实验者是对原假设 H_0: μ 的值接近于某个给定的 μ_0 及备择假设：μ 不接近于 μ_0 感兴趣. 在某些这类问题中，简单的假设 $H_0: \mu = \mu_0$ 只不过是对决策目标的理想化或简化. 在其他时候，应用更实际的复合原假设：指定 μ 属于 μ_0 附近的一个已知区间是有价值的，我们现在讨论这样的假设.

481

假定 $X_1, \cdots, X_n$ 属于服从正态分布的一个随机样本，且均值 μ 未知，方差 $\sigma^2 = 1$，假如要检验以下假设：

$$
\begin{aligned}
&H_0: 9.9 \leqslant \mu \leqslant 10.1, \\
&H_1: \mu < 9.9 \text{ 或 } \mu > 10.1.
\end{aligned}
\tag{8.2.4}
$$

由于备择假设 H_1 是双边的，所以再次看到，使用当 $\overline{X}_n \leqslant c_1$ 或 $\overline{X}_n \geqslant c_2$ 时拒绝 H_0 的检验是恰当的. 我们将确定当 $\mu = 9.9$ 或 $\mu = 10.1$ 时拒绝 H_0 的概率为 0.05 的 c_1 和 c_2 的值.

令 $\pi(\mu|\delta)$ 表示 δ 的功效函数. 当 $\mu = 9.9$，随机变量 $n^{1/2}(\overline{X}_n - 9.9)$ 服从标准正态分布. 因此

$$\begin{aligned}\pi(9.9|\delta) &= \Pr(\text{拒绝 } H_0 | \mu = 9.9) \\ &= \Pr(\overline{X}_n \leqslant c_1 | \mu = 9.9) + \Pr(\overline{X}_n \geqslant c_2 | \mu = 9.9) \\ &= \Phi[n^{1/2}(c_1 - 9.9)] + 1 - \Phi[n^{1/2}(c_2 - 9.9)].\end{aligned} \tag{8.2.5}$$

类似地，当 $\mu = 10.1$ 时，随机变量 $n^{1/2}(\overline{X}_n - 10.1)$ 服从标准正态分布且

$$\pi(10.1|\delta) = \Phi[n^{1/2}(c_1 - 10.1)] + 1 - \Phi[n^{1/2}(c_2 - 10.1)]. \tag{8.2.6}$$

必须令 $\pi(9.9|\delta)$ 与 $\pi(10.1|\delta)$ 都等于 0.05. 由于正态分布的对称性，如果 c_1 与 c_2 的取值关于 10 对称，那么功效函数 $\pi(\mu|\delta)$ 关于点 $\mu = 10$ 对称. 特别地，$\pi(9.9|\delta) = \pi(10.1|\delta)$ 成立.

相应地，取 $c_1 = 10 - c, c_2 = 10 + c$，则基于等式(8.2.5)和(8.2.6)，我们有

$$\pi(9.9|\delta) = \pi(10.1|\delta) = \Phi[n^{1/2}(0.1 - c)] + 1 - \Phi[n^{1/2}(0.1 + c)]. \tag{8.2.7}$$

c 值的确定要使 $\pi(9.9|\delta) = \pi(10.1|\delta) = 0.05$，所以，$c$ 的值要使得

$$\Phi[n^{1/2}(0.1 + c)] - \Phi[n^{1/2}(0.1 - c)] = 0.95. \tag{8.2.8}$$

对于任意给定的 n，符合等式(8.2.8)的 c 值都可通过反复尝试摸索从标准正态分布的表中找到.

比如说，当 $n = 16$，c 必须由

$$\Phi(0.4 + 4c) - \Phi(0.4 - 4c) = 0.95 \tag{8.2.9}$$

确定. 在尝试不同取值的 c 后，我们发现当 $c = 0.527$ 时等式(8.2.9)成立，因此

$$c_1 = 10 - 0.527 = 9.473 \text{ 和 } c_2 = 10 + 0.527 = 10.527.$$

482

因此，若 $n = 16$，当 $\overline{X}_n \leqslant 9.437$ 或 $\overline{X}_n \geqslant 10.527$ 时，过程 δ 拒绝 H_0，检验方法中的功效函数 $\pi(\mu|\delta)$，关于点 $\mu = 10$ 对称，且 $\pi(9.9|\delta) = \pi(10.1|\delta) = 0.05$，而且，当 $9.9 < \mu < 10.1$ 时，$\pi(\mu|\delta) < 0.05$，且当 $\mu < 9.9$ 或 $\mu > 10.1$ 时，$\pi(\mu|\delta) > 0.05$. 函数 $\pi(\mu|\delta)$ 如图 8-5 中所示.

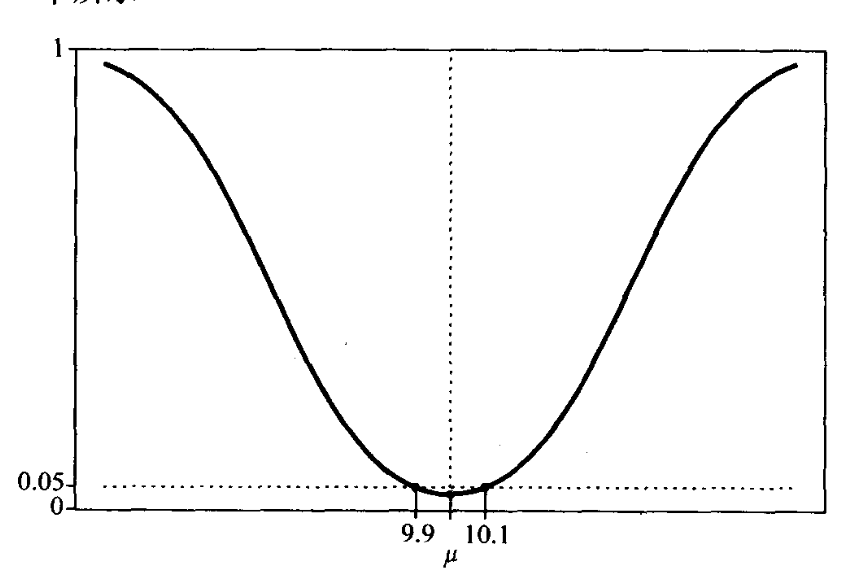


图 8-5 (8.2.4)中假设检验中的功效函数 $\pi(\mu|\delta)$

8.2.4　扩展研究：无偏检验

考虑检验以下假设的一般问题：

$$H_0: \theta \in \Omega_0,$$
$$H_1: \theta \in \Omega_1.$$

照例，设$\pi(\theta|\delta)$表示检验δ的功效函数. 如果对于任意一对参数值$\theta \in \Omega_0$以及$\theta' \in \Omega_1$，有

$$\pi(\theta|\delta) \leqslant \pi(\theta'|\delta), \tag{8.2.10}$$

则称该检验δ是无偏的. 换言之，当$\theta \in \Omega_1$的功效函数的值至少与$\theta \in \Omega_0$的功效函数的值一样大，则δ是无偏检验.

无偏检验的概念是很吸引人的. 因为假设检验的目标是当$\theta \in \Omega_0$时接受H_0，而当$\theta \in \Omega_1$就拒绝H_0，所以希望只要当$\theta \in \Omega_0$，拒绝H_0的概率至少与$\theta \in \Omega_1$的概率一样大. 可以看出功效函数如图8-5所示的检验δ是假设(8.2.4)的无偏检验. 再看图8-4，在检验假设(8.2.1)时，该图中的四种功效函数所给的四种检验中，只有δ_3是无偏的. 尽管这超出了本书的范围，但人们可以证明在所有水平为$\alpha_0 = 0.05$下的无偏检验中，δ_3是一致最大功效检验(UMP).

483

使得一个检验成为无偏检验的条件有时可以缩小检验的选择范围. 但在比较特殊的环境下，我们才去寻求无偏检验. 比如，统计学家在检验假设(8.2.4)的时候，只要他认为对于任意$a>0$，当$\theta=10.1+a$时拒绝H_0与当$\theta=9.9-a$时拒绝H_0是一样重要的，同时他也认为这两个θ值的发生是等可能的时候，他才应该使用如图8-5中显示的无偏检验δ. 在实践中，如果一个有偏检验的功效函数在Ω_1的特定区域里有较大的值，而这些特定区域他认为是特别重要的，或者非常可能包含θ的真值，则他更愿意用有偏检验而不用无偏检验.

在本章的剩余部分里，我们要考虑的是，在实际运用中常常遇到的特殊检验情况. 在每一个情况中，对于每个α_0，不存在水平为α_0的一致最大功效检验，我们将导出在这些情况下的检验，它们以似然比为基础. 但在高级课程中，会说明在8.3，8.4和8.5节给出的t检验与F检验在具有它们的水平的各种无偏检验类中，是UMP的.

8.2.5　小结

假定已知方差的正态分布的均值是特定值，而备择假设是双侧的，在这样的情况下，结合两个检验水平为α_1和α_2的单边检验的拒绝域，取水平$\alpha_0=\alpha_1+\alpha_2$，一般人们喜欢取$\alpha_1=\alpha_2=\alpha_0/2$，在这种情况下，若一个随机样本$X_1,\cdots,X_n$取自于均值为$\mu$、方差为$\sigma^2$的正态分布，可以检验：$H_0:\mu=\mu_0$和$H_1:\mu\neq\mu_0$，当$\overline{X}_n>\mu_0+\Phi^{-1}(1-\alpha_0/2)\sigma/n^{1/2}$或当$\overline{X}_n<\mu_0-\Phi^{-1}(1-\alpha_0/2)\sigma/n^{1/2}$，就拒绝原假设$H_0$. 这里$\Phi^{-1}$是标准正态分布的分位数函数，一个检验的功效函数在备择假设点上的值比在原假设每个

点上的值大，那么该检验是无偏检验. 上面所说的 $\alpha_1=\alpha_2=\alpha_0/2$ 的正态分布检验就是无偏的.

习题

1. 假定 $X_1,\cdots,X_n$ 是服从均值 μ 未知、方差为 1 的正态分布的随机样本. 要求对于给定的 μ_0 检验以下假设：
$$H_0:\mu=\mu_0,$$
$$H_1:\mu\neq\mu_0.$$
考虑一个检验过程 δ，若 $\overline{X}_n\leqslant c_1$ 或 $\overline{X}_n\geqslant c_2$，则拒绝原假设 H_0. 设 $\pi(\theta|\delta)$ 表示 δ 的功率函数，确定常数 c_1，c_2 的值，使得 $\pi(\mu_0|\delta)=0.10$ 并且函数 $\pi(\mu|\delta)$ 关于 $\mu=\mu_0$ 对称.
2. 再次考虑习题 1 的条件，并假定
$$c_1=\mu_0-1.96n^{-1/2}.$$
确定当 $\pi(\mu_0|\delta)=0.10$ 时 c_2 的值.
3. 再次考虑习题 1 的条件和该题描述的检验步骤. 当 $\pi(\mu_0|\delta)=0.10$ 并且 $\pi(\mu_0+1|\delta)=\pi(\mu_0-1|\delta)\geqslant0.95$ 时，求 n 的最小值.

484

4. 设 $X_1,\cdots,X_n$ 是服从于均值 μ 未知、方差为 1 的正态分布的随机样本. 要求检验如下假设：
$$H_0:0.1\leqslant\mu\leqslant0.2,$$
$$H_1:\mu<0.1 \text{ 或 } \mu>0.2.$$
考虑一个检验过程 δ，若 $\overline{X}_n\leqslant c_1$ 或 $\overline{X}_n\geqslant c_2$，则拒绝原假设 H_0. 设 $\pi(\mu|\delta)$ 表示 δ 的功率函数. 设样本容量 $n=25$，确定当 $\pi(0.1|\delta)=\pi(0.2|\delta)=0.07$ 时常量 c_1，c_2 的值.
5. 再次考虑习题 4 的条件，并且设 n 也等于 25，确定当 $\pi(0.1|\delta)=0.02$ 和 $\pi(0.2|\delta)=0.05$ 时，常量 c_1,c_2 的值.
6. 设随机样本 $X_1,\cdots,X_n$ 取自区间 $[0,\theta]$ 上的均匀分布，θ 的值未知，要求检验如下假设：
$$H_0:\theta\leqslant3,$$
$$H_1:\theta>3.$$
   a. 对于每一个显著性水平 $\alpha_0(0\leqslant\alpha_0<1)$，存在一个一致最大功效检验，它规定当 $\max(X_1,\cdots,X_n)\geqslant c$ 时拒绝 H_0；
   b. 对于每一个可能的 α_0 值，确定 c 的值.
7. 对于一个给定的样本容量 n 和一个给定的 α_0 值，作出习题 6 中一致最大功效检验的功效函数的草图.
8. 设 $X_1,\cdots,X_n$ 是取自于习题 6 所述的均匀分布的随机样本，但假定现在欲检验如下的假设：
$$H_0:\theta\geqslant3,$$
$$H_1:\theta<3.$$
   a. 对于每个显著性水平 $\alpha_0(0\leqslant\alpha_0<1)$，存在一个一致最大功效检验，当 $\max(X_1,\cdots,X_n)\leqslant c$ 时拒绝 H_0；
   b. 对于每一个可能的 α_0 值，确定 c 的值.
9. 对于一个给定的样本容量 n 和一个给定的 α_0 值，给出习题 8 中得到的一致最大功效检验的功效函数.

10. 设 $X_1,\cdots,X_n$ 是取自于习题 6 所述均匀分布的随机样本，但假定现在欲检验如下的假设：

$$H_0: \theta=3,$$
$$H_1: \theta\neq 3.$$

考虑一个检验过程 δ，若 $\max(X_1,\cdots,X_n)\leqslant c_1$ 或 $\max(X_1,\cdots,X_n)\geqslant c_2$，则拒绝原假设 H_0. 设 $\pi(\theta|\delta)$ 表示 δ 的功效函数. 确定当 $\pi(3|\delta)=0.05$ 且 δ 为无偏时常量 c_1，c_2 的值.

11. 再次考虑习题 1 的条件，确定当 $\pi(\mu_0|\delta)=0.10$ 且 δ 为无偏时常量 c_1，c_2 的值.

12. 令 X 服从参数为 β 的指数分布，假定希望检验以下假设：

$$H_0: \beta=1,$$
$$H_1: \beta\neq 1.$$

将采用如下的检验程序，当 $X\leqslant c_1$ 或 $X\geqslant c_2$ 时，拒绝原假设 H_0 的检验.

a. 构建 c_1 和 c_2 必须满足的方程，使得检验过程的显著性水平为 α_0.

b. 求一对有限的、非零的值 (c_1,c_2)，使得检验过程的显著性水平为 $\alpha_0=0.1$.

8.3 t 检验

我们从处理检验正态分布参数假设的几种特殊情形入手. 在本节中，将解决方差和均值都未知的情况下的问题，讨论关于平均值的假设检验，此类检验都以 t 分布为基础.

485

8.3.1 检验当方差未知时关于正态分布均值的假设

在本部分，将考虑均值和方差都未知时，检验正态分布均值的假设问题. 明确地说，设变量 $X_1,\cdots,X_n$ 是取自均值 μ 和方差 σ^2 都未知的正态分布的随机样本，考虑检验如下的假设：

$$H_0: \mu\leqslant\mu_0,\qquad (8.3.1)$$
$$H_1: \mu>\mu_0.$$

在该问题中参数空间 Ω 包含了所有二维向量 (μ,σ^2)，其中 $-\infty<\mu<\infty$ 且 $\sigma^2>0$. 原假设 H_0 指定了向量 (μ,σ^2) 落在 Ω 的子集 Ω_0 部分，Ω_0 包含所有 $\mu\leqslant\mu_0$ 且 $\sigma^2>0$ 的向量，如图 8-6 所示. 备择假设 H_1 指定 (μ,σ^2) 位于 Ω 中的 Ω_1 部分，Ω_1 由所有不属于 Ω_0 的向量组成.

在例 8.1.10 中，已经演示了如何从 μ 的单侧置信区间中得到 (8.3.1) 的检验假设. 具体来说，定义 $\overline{X}_n=\sum_{i=1}^n X_i/n$，$\sigma'=\left(\sum_{i=1}^n (X_i-\overline{X}_n)^2/[n-1]\right)^{1/2}$，且

$$U=n^{1/2}\ \frac{\overline{X}_n-\mu_0}{\sigma'}.\qquad (8.3.2)$$

若 $U\geqslant c$ 则拒绝 H_0. 当 $\mu=\mu_0$ 时，由第 7.4 节可以得出，无论 σ^2 为何值，由 (8.3.2) 规定的统计量 U 将服从自由度为 $n-1$ 的 t 分布. 由于该原因，基于 U 的检验称为 t 检验.

486

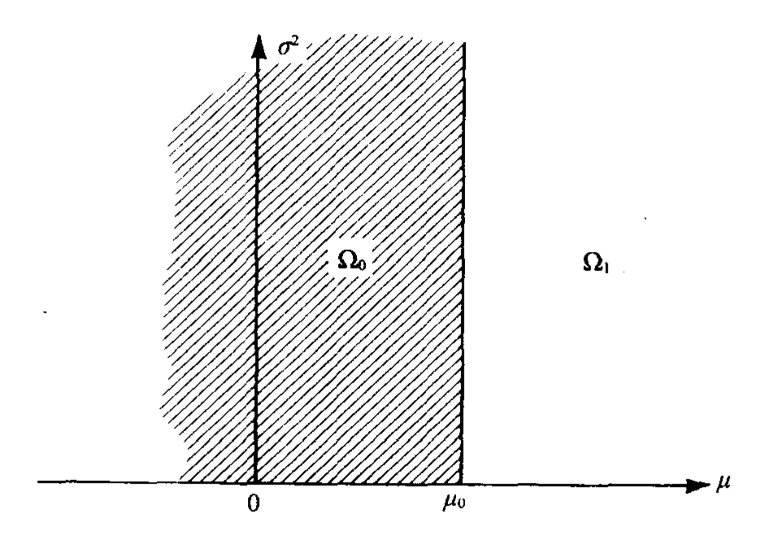


图 8-6 假设(8.3.1)中参数空间 Ω 的子集 Ω_0 和 Ω_1

8.3.2 t 检验的显著性水平

由于 U 服从自由度为 $n-1$ 的 t 分布，若 $U \geq T_{n-1}^{-1}(1-\alpha_0)$，即自由度为 $n-1$ 的 t 分布的 $1-\alpha_0$ 分位数，要拒绝原假设 H_0. 现在来证明该检验是 α_0 水平的检验.

设 $\mu < \mu_0$，$c = T_{n-1}^{-1}(1-\alpha_0)$，再令

$$U^* = \frac{n^{1/2}(\overline{X}_n - \mu)}{\sigma'}, \quad W = \frac{n^{1/2}(\mu_0 - \mu)}{\sigma'}.$$

则随机变量 U^* 服从自由度为 $n-1$ 的 t 分布，且随机变量 W 的值为正. 此外，由式(8.3.2)可知 $U = U^* - W$. 所以，对于 $\mu < \mu_0$ 的所有的 μ 和 σ^2 的值，有：

$$\Pr(U \geq c) = \Pr(U^* - W \geq c) = \Pr(U^* \geq c + W) < \Pr(U^* \geq c) = \alpha_0.$$

由此，当 $\mu < \mu_0$ 时，无论 σ^2 取何值，都有 $\Pr(U \geq c) < \alpha_0$. 此时的 α_0 就是在 $U \geq c$ 时拒绝 H_0 的检验水平.

例 8.3.1 纤维的长度 现有一批生产加工出来的金属纤维，设他们的长度(单位：毫米)服从正态分布，均值 μ 与方差 σ^2 未知. 考虑检验如下假设：

$$H_0: \mu \leq 5.2, \quad H_1: \mu > 5.2. \tag{8.3.3}$$

现已从这批纤维中随机选取了 15 根测量其长度，得样本均值 $\overline{X}_{15} = 5.4$，$\sigma' = 0.4226$. 在这些数据的基础上，我们来进行在显著性水平 $\alpha_0 = 0.05$ 下的 t 检验.

由 $n = 15$，$\mu_0 = 5.2$，由等式(8.3.2)定义的统计量 U 在 $\mu = 5.2$ 时应服从自由度为 14 的 t 分布. 查 t 分布表可得 $T_{14}^{-1}(0.95) = 1.761$. 所以，当 $U > 1.761$ 时，原假设 H_0 将被拒绝. 由等式(8.3.2)可计算出 U 的值为 1.833. 所以在 0.05 的显著性水平下拒绝 H_0. ◀

由假设(8.3.1)和统计量 U 的一个观测值，可以如下计算出 p-值. 首先注意到在

$U \geqslant c$时拒绝 H_0. 当 $U=u$ 时，p-值可根据等式(8.1.10)来计算得出. 等式(8.1.10)的上确界在$\mu \leqslant \mu_0$ 时的所有有序数对(μ,σ^2)之上. 所以，p-值为：

$$\sup_{\mu \leqslant \mu_0, \sigma^2} \Pr(U \geqslant c \mid \mu,\sigma^2) = \Pr(U \geqslant c \mid \mu_0,\sigma^2) = 1 - T_{n-1}(u).$$

此处的第一个等式已在本节早些时候给出了证明. 在例 8.3.1 中，已知 $u=1.833$，$n=15$，可以用计算机软件计算出 $1-T_{14}(1.833)=0.0441$. 487

8.3.3 t检验的功效

t检验的功效函数有以下的5个性质：

i. $\mu=\mu_0$ 时，$\pi(\mu,\sigma^2 \mid \delta)=\alpha_0$，　　ii. $\mu<\mu_0$ 时，$\pi(\mu,\sigma^2 \mid \delta)<\alpha_0$，

iii. $\mu>\mu_0$ 时，$\pi(\mu,\sigma^2 \mid \delta)>\alpha_0$，　　iv. $\mu \to -\infty$，$\pi(\mu,\sigma^2 \mid \delta) \to 0$，

v. $\mu \to \infty$，$\pi(\mu,\sigma^2 \mid \delta) \to 1$.

与证明 t 检验的显著性水平为 α_0 相类似，可以证明：若$\mu<\mu_0$，则无论 σ^2 取何值，都有 $\Pr(U \geqslant c)>\alpha_0$.

最后的两点性质证明起来有些困难. 当μ 值很大时，U 的分子部分趋向于极大，此时拒绝 H_0 的概率将接近于1. 正式的证明可以得到：对于每个 $\sigma^2>0$，

$$\lim_{\mu \to \infty} \pi(\mu,\sigma^2 \mid \delta)=1.$$

类似地，可以证明：对于每个 $\sigma^2>0$，

$$\lim_{\mu \to -\infty} \pi(\mu,\sigma^2 \mid \delta)=0.$$

如果知道等式(8.3.2)定义的统计量 U 的分布，那么对于所有的μ 值就能确定此时 t 检验的功效函数了. 可以把 U 改写成：

$$U=\frac{n^{1/2}(\overline{X}_n-\mu_0)/\sigma}{\sigma'/\sigma}. \tag{8.3.4}$$

等式(8.3.4)右边的分子部分服从均值为 $n^{1/2}(\mu-\mu_0)/\sigma$，方差为1的正态分布. 分母部分是一个$\chi^2$ 随机变量除以他的自由度 $n-1$ 所得商的平方根. 要不是因为非零均值的缘故，如同已经证明过的那样，该比值本该服从 t 分布. 当分子部分的均值不为0时，U 服从非中心 t 分布.

1. 非中心 t-分布

如果 Y 和 W 是独立的随机变量，其中 W 服从均值为 ψ 且方差为1的正态分布，Y 服从自由度为 m 的χ^2 分布，则变量

$$X=\frac{W}{\left(\frac{Y}{m}\right)^{1/2}}$$

服从的分布称为自由度为 m 非中心参数为 ψ 的非中心 t 分布. 令 $T_m(t \mid \psi)$表示该分布的分布函数，即 $T_m(t \mid \psi)=\Pr(X \leqslant t)$（如果 $\psi=0$，这样的非中心 t 分布也是一个 t 分布）. 于是，对于给定了的μ，U 的分布是自由度为 $n-1$，非中心参数为 $\psi=n^{1/2}(\mu-\mu_0)/\sigma$ 的非中心 t 分布. 取 δ 使得当 $U \geqslant c$ 时，拒绝原假设 H_0: $\mu \leqslant \mu_0$，则 δ 的功效 488

函数为$\pi(\mu,\sigma^2|\delta)=1-T_{n-1}(c|\psi)$. 在习题 11 中，你可以证明公式 $1-T_m(t|\psi)=T_m(-t|-\psi)$. 有些计算机程序可以用来计算非中心 t 分布的分布函数，一些统计软件包也包含了这样的程序. 图 8-7 给出了在置信水平为 0.05 和 0.01 时，不同自由度和不同非中心参数值时 t 检验的功效函数. 我们用横轴表示 $|\psi|$，因为同样的图形可用来表示两种类型的单边假设. 接下来的例子描绘了怎样用图 8-7 来近似得到所要求的功效.

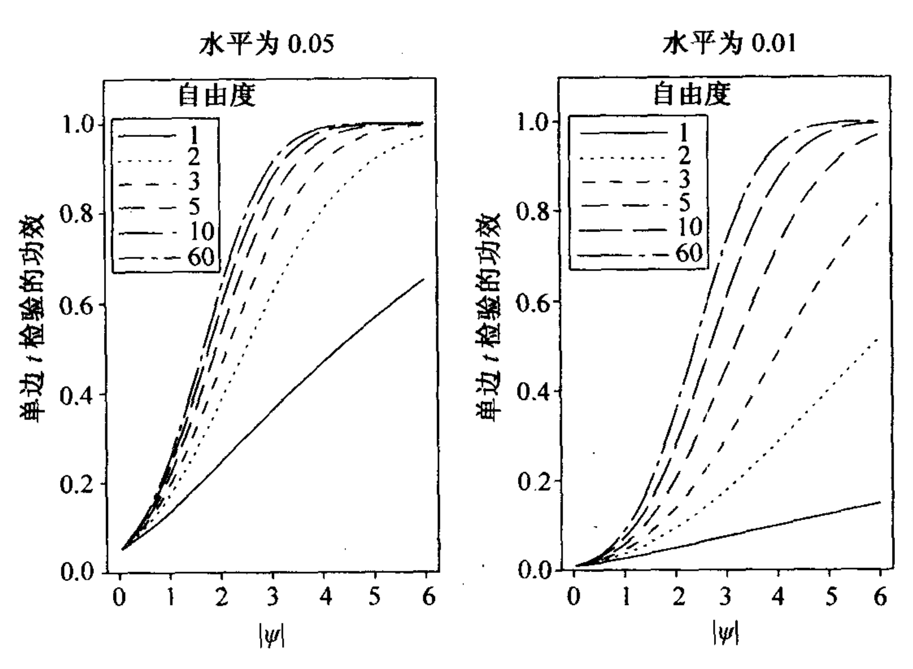


图 8-7 不确定参数 ψ 的各种取值和各种不同自由度时在水平 0.05 和 0.01 时单边 t 检验的功效函数

例 8.3.2 光纤的长度 在例 8.3.1 中，我们在置信水平为 0.05 上检验了假设 (8.3.3). 假定当 μ 不等于 5.2 的时候，对检验的功效感兴趣. 特别地，假定当 $\mu=5.2+\sigma/2$，即 5.2 加上标准差的一半，我们要得到相应的功效，则非中心参数

$$\psi=15^{1/2}\left(\frac{5.2+\sigma/2-5.2}{\sigma}\right)=1.936.$$

在图 8-7 中，没有自由度为 14 的曲线，然而自由度为 10 和 60 的曲线之间并没有太大的差别. 因此我们可以认为所要的答案就在这两者之间，如果在图 8-7 中观察置信水平为 0.05 的图，并且从横轴上的 1.936(大约为 2)开始往上找，直到当超过自由度为 10 的那条曲线. 此时我们发现功效大约为 0.6(实际功效为 0.578). ◀

注：功效是非中心参数的函数. 在例 8.3.2 中，不能回答类似于"当 $\mu=5.5$ 时置信水平为 0.05 的检验的功效是什么"的问题. 原因是功效通过非中心参数而成为 μ 和 σ^2 共同的函数（见习题6）. 对于每一个可能的 σ^2 和 $\mu=5.5$，非中心参数 $\psi=15^{1/2}\times 0.3/\sigma$，$\psi$ 从 0 到 ∞ 的变化取决于 σ. 这就是为什么，每当我们想得到一个 t 检验功效的数值时，总需要知道 μ 和 σ 的取值，或者知道 μ 离开 μ_0 是 σ 的几倍.

489

如果我们想要在置信水平 α_0 下检验假设

$$H_0:\mu\geqslant\mu_0,\qquad (8.3.5)$$
$$H_1:\mu<\mu_0.$$

那么如果 $U<-c, c=T_{n-1}^{-1}(1-\alpha_0)$，则拒绝 H_0. 该检验 δ'的功效函数为

$$\pi(\mu,\sigma^2|\delta')=T_{n-1}(-c|\psi),$$

其中 $\psi=n^{1/2}(\mu-\mu_0)/\sigma<0$. 在习题11中，可以证明该功效等于 $1-T_{n-1}(c|-\psi)$. 这个值在例8.3.2的图8-7中可以找到. 在图8-7中的横轴上找到 $|\psi|$，然后再沿垂直方向找到自由度接近的曲线. 为了确定 p-值，首先注意到当 $-U>-c$ 时，拒绝 H_0. $-U$ 的分布在 $\mu=\mu_0$ 时所有参数仍然是自由度为 $n-1$ 的 t 分布. 如果 $-U=-u$，则 p-值就是 $1-T_{n-1}(-u)=T_{n-1}(u)$.

2. 选择样本容量大小

利用功效函数来帮助确定合适的样本容量是可能的. 比如，在例8.3.2中，当 $\mu=5.2+\sigma/2$ 时得出了功效. 假定在 $\mu=5.2+\sigma/2$ 的时候，要求得到的功效接近于0.8，而不是0.6，就会用多于15个的观测值来达到这样的效果. 在图8-7中，可以看到非中心参数 ψ 的何种数值可以使得功效达到0.8. 若自由度在10和60之间，需要大约为2.5的 ψ. 但是当 $\mu=5.2+\sigma/2$ 时，$\psi=n^{1/2}/2$. 因此，需要的样本容量 n 大约为25. 精确的计算表明，当 $n=25$，$\mu=5.2+\sigma/2$ 的时候，检验水平为0.05的功效为0.7834. 当 $n=26$ 的时候，功效为0.7981；而 $n=27$ 的时候，功效为0.8118.

8.3.4 成对的 t 检验

在很多实验中，在相同的实验装置下使用两种不同条件测量同一个变量，我们感兴趣的是在一种条件下的均值是不是比在另一种条件下的大？在这样的情况下，通常是将两组测量值对应相减，再将所得差值看作是服从正态分布的随机变量. 然后再做关于差值的均值的假设检验.

例8.3.3 假人撞击实验 国家交通运输安全委员会通过对放在车内的假人在汽车受到撞击后的受伤部位及伤害程度的实验中收集到一些数据. 在一系列的实验中，在每辆车上的驾驶席上放置一个假人，在副驾驶席上放置另一个假人. 要测量的一个变量是每个假人头部受伤的程度. 图8-8是不同位置的假人头部受伤数据的双对数坐标图. 我们关注的问题之一是在驾驶席和副驾驶席上的假人的头部受伤程度是否有差异，以及或者是二者之间的差异程度. 令 $X_1,\cdots,X_n$ 是驾驶席与副驾驶席的假人的头部受伤程度的对数之差. 可以假定 $X_1,\cdots,X_n$ 是服从均值为 μ，方差为 σ^2 的正态分布的随机变量. 如果希望在 $\alpha_0=0.01$ 的显著性水平下检验相对于备择假设为 $\mu>0$的原假设 $\mu\leqslant0$. 图8-8中显示的是 $n=164$ 辆车的数据. 如果 $U\geqslant T_{163}^{-1}(0.99)=2.35$，则检验拒绝原假设 H_0.

490

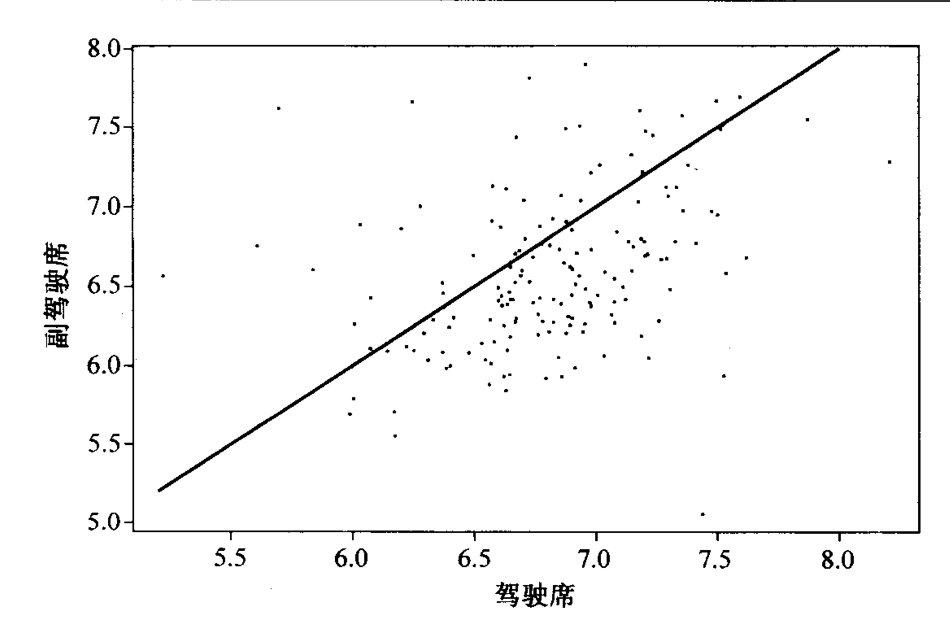


图 8-8 驾驶席和副驾驶席假人头部受伤数据的对数坐标图. 直线表示两个数据相等

图 8-8 中坐标值之差的均值为 $\bar{x}_n = 0.2199$，方差 $\sigma' = 0.5342$，于是 U 值为 5.271，大于 2.35，因此在 0.01 的显著性水平下拒绝原假设. 事实上，p-值是小于 10^{-6}.

如果还想确定在显著性水平 0.01 下备择假设 H_1 下的功效函数. 假定驾驶席与副驾驶席的假人头部受伤值的对数之平均值差为 $\sigma/4$，则非中心参数为 $(164)^{1/2}/4 = 3.20$. 在图 8-7 的右边的图表中，功效值大约为 0.8.（实际上，是 0.802.）——◀

8.3.5 备择假设为双侧的假设检验

仍然设变量 $X_1, \cdots, X_n$ 组成的服从均值为 μ、方差为 σ^2 的正态分布的随机样本，其中 μ 和 σ^2 均为未知，但现在要对下面的假设进行检验：

$$H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0. \tag{8.3.6}$$

这里，备择假设 H_1 是双边的.

491

在例 8.1.8 中，从 7.5 节讨论的置信区间出发，导出了假设(8.3.6)的 α_0 水平检验. 该检验中有“如果 $|U| \geq T_{n-1}^{-1}(1-\alpha_0/2)$，则拒绝原假设 H_0”的形式，其中 T_{n-1}^{-1} 是自由度为 $n-1$ 的 t 分布的分位数函数，U 由式(8.3.2)定义. 我们知道，对于满足原假设 H_0 的 μ 和 σ^2 的所有取值，即当 $\mu = \mu_0$ 且 $\sigma^2 > 0$ 时，统计量 U 服从自由度为 $n-1$ 的 t 分布.

例 8.3.4 纤维的长度 重新考虑在例 8.3.1 中所讨论过的问题，但是现在代替假设(8.3.3)的是如下的假设：

$$H_0: \mu = 5.2, \quad H_1: \mu \neq 5.2. \tag{8.3.7}$$

仍假定测量了15根纤维的长度，从观测值计算的 U 值为1.8333，在显著性水平 $\alpha=0.05$ 下检验假设.

因为 $\alpha=0.05$ 的临界值是自由度为14的 t 分布的 $1-0.05/2=0.975$ 分位数. 从本书的 t 分布表，找到 $T_{14}^{-1}(0.975)=2.145$，所以 t 检验将在 $U<-2.145$ 或 $U>2.145$ 时拒绝 H_0. 因为 $\mu=1.833$，所以接受原假设 H_0. ◀

在给定的问题中，确定适当的备择假设是单侧的还是双侧的是很重要的，例8.3.1和例8.3.4中的数值强调了这一点. 当在显著性水平0.05下检验假设(8.3.3)，原假设 $H_0:\mu\leqslant5.2$ 被拒绝. 当在相同显著性水平下检验假设(8.3.7)时，使用了相同数据，原假设 $H_0:\mu=5.2$ 被接受.

1. 双侧检验的功效函数

当 $|U|>c$，$c=T_{n-1}(1-\alpha_0/2)$ 时拒绝 $H_0:\mu=\mu_0$ 的检验 δ 的功效函数可以通过使用非中心 t 分布得到. 如果 $\mu\neq\mu_0$，那么 μ 服从一个自由度为 $n-1$、非中心参数为 $\psi=n^{1/2}(\mu-\mu_0)/\sigma$ 的 t 分布，和我们在单侧假设检验中得到的一样. 于是 δ 的功效函数为

$$\pi(\mu,\sigma^2|\delta)=T_{n-1}(-c|\psi)+1-T_{n-1}(c|\psi).$$

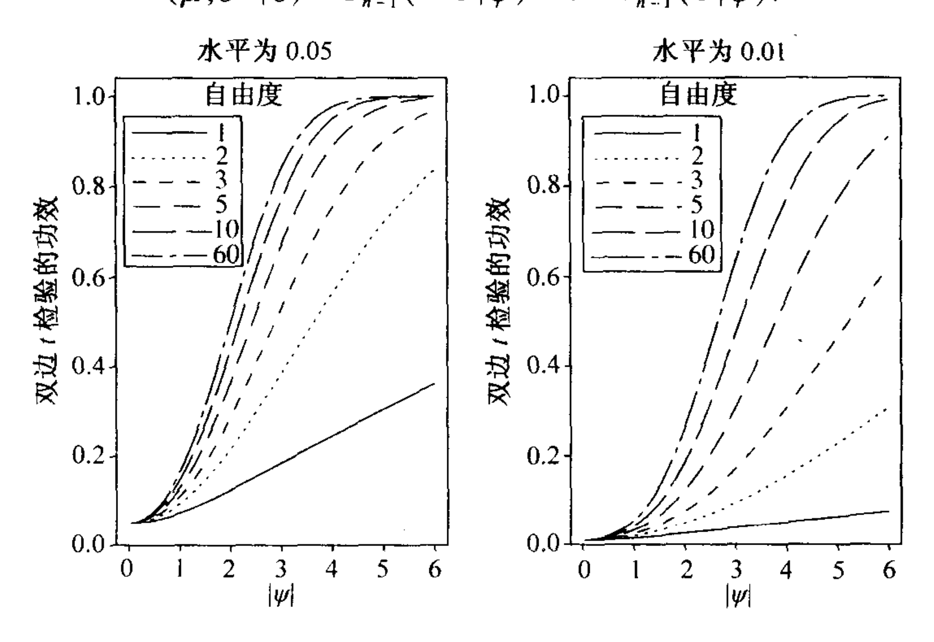


图8-9 不确定参数 ψ 的各种取值和各种不同自由度时在水平0.05和0.01时双边 t 检验的功效函数

图8-9画出了不同自由度和非中心参数的功效函数. 可以用图8-9找到例8.3.4当 $\mu=5.2+\sigma/2$，即 $\psi=1.936$ 检验的功效大约是0.45.（确切的功效值为0.438).

2. 双侧检验的 p-值

当 $U=u$ 时，可以如下计算假设(8.3.6)的 p-值. 当 $|U|\geqslant c$ 时拒绝 H_0，当 $\mu=\mu_0$ 时，$|U|$ 的分布函数为 $F_{\mu_0}(|u|)=2T_{n-1}(|u|)-1$. 所以 p-值是 $1-F_{\mu_0}(|u|)=2[1-T_{n-1}(|u|)]$. 例如，在例8.3.4中，$p$-值为 $2[1-T_{14}(1.833)]=0.0882$. 注意，

这是当假设是(8.3.1)时 p-值的两倍. 就 t 检验来说，如果(8.3.1)或(8.3.5)假设的 p-值为 p，那么(8.3.6)假设的 p-值就是 $2p$ 或 $2(1-p)$.

492

8.3.6 扩展研究：似然比检验

一个非常受欢迎的假设检验是似然比检验. 它基于似然函数 $f_n(\boldsymbol{x}\mid\theta)$. 似然函数往往当接近 θ 的真实值时趋于最大. 这就是极大似然估计在许多情况下都起到很大作用的原因. 现在假定我们想要检验假设

$$H_0: \theta\in\Omega_0,\qquad H_1: \theta\in\Omega_1. \tag{8.3.8}$$

为了比较这两个假设，也许希望看看似然函数在 Ω_0 上更大还是在 Ω_1 上更大. 比如说，如果只限于关注 H_0，可以计算极大似然估计值 $\sup_{\theta\in\Omega_0} f_n(\boldsymbol{x}\mid\theta)$. 类似地，对 H_1 可以计算 $\sup_{\theta\in\Omega_1} f_n(\boldsymbol{x}\mid\theta)$，这两个值的比值可以用于检验假设(8.3.8)，定义似然比统计量

$$\Lambda(x)=\frac{\sup_{\theta\in\Omega_1} f_n(\boldsymbol{x}\mid\theta)}{\sup_{\theta\in\Omega_0} f_n(\boldsymbol{x}\mid\theta)}. \tag{8.3.9}$$

那么，假设(8.3.8)的似然比检验，就是对于某一常数 k，如果 $\Lambda(\boldsymbol{x})\geqslant k$，则拒绝 H_0，即，如果 Ω_1 的似然函数相对于 Ω_0 的似然函数来说足够大时拒绝 H_0. 通常，如果可能的话，k 的选择应使检验有期望的水平 α_0.

493

例如，考虑假设(8.3.1). 在观察到随机样本值 $x_1,\cdots,x_n$ 以后，似然函数为

$$f_n(\boldsymbol{x}\mid\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right], \tag{8.3.10}$$

其中 $\Omega_0=\{(\mu,\sigma^2)\mid\mu\leqslant\mu_0\}$且 $\Omega_1=\{(\mu,\sigma^2)\mid\mu>\mu_0\}$，似然比统计量为

$$\Lambda(\boldsymbol{x})=\frac{\sup_{\{(\mu,\sigma^2)\mid\mu>\mu_0\}} f_n(\boldsymbol{x}\mid\mu,\sigma^2)}{\sup_{\{(\mu,\sigma^2)\mid\mu\leqslant\mu_0\}} f_n(\boldsymbol{x}\mid\mu,\sigma^2)}. \tag{8.3.11}$$

现在来推导基于(8.3.11)的似然比检验的显式表达式，如6.4节那样，当只知道点(μ,σ^2)属于参数空间 Ω 时，用 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 代表 μ 和 σ^2 的极大似然估计. 在例6.4.4中已经证明了

$$\hat{\mu}=\bar{x}_n \text{ 且 } \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2.$$

同理，当点(μ,σ^2)限制在子集 Ω_0 内时，用 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 代表 μ 和 σ^2 的 M. L. E.. 当点(μ,σ^2)落在子集 Ω_1 中时，用 $\hat{\mu}_1$ 和 $\hat{\sigma}_1^2$ 代表 μ 和 σ^2 的 M. L. E.. 于是，

$$\Lambda(x)=\frac{f_n(\boldsymbol{x}\mid\hat{\mu}_1,\hat{\sigma}_1^2)}{f_n(\boldsymbol{x}\mid\hat{\mu}_0,\hat{\sigma}_0^2)}. \tag{8.3.12}$$

首先假定所取的样本值使得 $\bar{x}_n>\mu_0$. 为使 $\hat{\mu}_1=\hat{\mu}$ 及 $\hat{\sigma}_1^2=\hat{\sigma}^2$，点$(\hat{\mu},\sigma^2)$应在 Ω_1 中.

此外，当 $\bar{x}_n > \mu_0$ 时，如果 μ 尽可能地靠近于 $\bar{x}_n$，$f_n(\boldsymbol{x}|\mu,\sigma^2)$ 将达到在所有 $(\mu,\sigma^2)\in\Omega_0$ 时的最大值，在子集 Ω_0 的所有点 μ 中最接近于 $\bar{x}_n$ 的值为 $\mu=\mu_0$，所以，$\hat{\mu}_0=\mu_0$. 像在例 6.4.4 中所做的那样，可以转而证明 σ^2 的极大似然估计为：

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_0)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2.$$

将刚才得到的 $\hat{\mu}_0$，$\hat{\sigma}_0^2$，$\hat{\mu}_1$ 和 $\hat{\sigma}_1^2$ 值代入方程(8.3.12)，发现：

$$\Lambda(\boldsymbol{x}) = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{n/2}. \tag{8.3.13}$$

如果用关系式：

494

$$\sum_{i=1}^{n}(x_i-\mu_0)^2 = \sum_{i=1}^{n}(x_i-\bar{x}_n)^2 + n(\bar{x}_n-\mu_0)^2,$$

则 $\Lambda(\boldsymbol{x})$ 就能被表示为：

$$\Lambda(\boldsymbol{x}) = \left[1+\frac{n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}\right]^{n/2}. \tag{8.3.14}$$

现在考虑任意指定的常数 $k>1$，由方程(8.3.14)可知，当且仅当

$$\frac{n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2} \geqslant k' \tag{8.3.15}$$

时有 $\Lambda(\boldsymbol{x})\geqslant k$，此处的 k' 是一个源自 k 的另一个常数. 最后，当 $\bar{x}_n>\mu_0$，关系式(8.3.15)等价于

$$\frac{n^{1/2}(\bar{x}_n-\mu_0)}{\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}{n-1}\right]^{1/2}} \geqslant c, \tag{8.3.16}$$

此时 $c=[(n-1)k']^{1/2}$.

同理，如果取定的样本值使得 $\bar{x}_n<\mu_0$，且如果 $k<1$，那么，可以再次证明 $\Lambda(\boldsymbol{x})\geqslant k$ 当且仅当关系式(8.3.16)成立，其中 c 是常数，它的值可以从 k 的值得到. 注意(8.3.16)左边的数值，是从(8.3.2)式得到的统计量 U 的观察值，所以该 t 检验也是似然比检验. 不难把上面的讨论作一些变动从而找到假设(8.3.5)和(8.3.6)的似然比检验(比如,参见习题 17 和 18).

注： 似然比的另外定义. 许多作者定义似然比检验的统计量为：

$$\Lambda'(\boldsymbol{x}) = \frac{\sup_{\theta\in\Omega_0} f_n(\boldsymbol{x}|\theta)}{\sup_{\theta\in\Omega} f_n(\boldsymbol{x}|\theta)}. \tag{8.3.17}$$

于是，似然比检验的过程就变为 $\Lambda'(\boldsymbol{x})\leqslant k'$ 时拒绝 H_0，此处 k' 是常数. 当 $k'<1$ 时，很容易证明：该似然比检验与所原先定义的似然比检验等价（见习题 19).

8.3.7 小结

若 $X_1,\cdots,X_n$ 组成一组服从均值 μ 及方差 σ^2 都未知的正态分布的随机样本，可以利用 $n^{1/2}(\overline{X}_n-\mu_0)/\sigma'$ 服从自由度为 $n-1$ 的 t 分布的事实做有关 μ 的假设检验. 用 T_{n-1}^{-1} 表示自由度为 $n-1$ 的 t 分布的分位数函数. 例如在检验水平 α_0 下要检验 $H_0:\mu\leqslant\mu_0$，相对地，$H_0:\mu>\mu_0$. 当 $n^{1/2}(\overline{X}_n-\mu_0)/\sigma'>T_{n-1}^{-1}(1-\alpha_0)$ 时，拒绝原假设 H_0. 要检验 $H_0:\mu=\mu_0$ 相对于 $H_1:\mu\neq\mu_0$. 当 $|n^{1/2}(\overline{X}_n-\mu_0)/\sigma'|\geqslant T_{n-1}^{-1}(1-\alpha_0/2)$ 时，拒绝原假设 H_0. 每一个检验的功效函数都可以用自由度为 $n-1$、非中心参数为 $\psi=n^{1/2}(\mu-\mu_0)/\sigma$ 的非中心 t 分布的密度函数来表示. 495

习题

1. 利用例 7.5.2 的数据构成奶酪中乳酸量的 $n=10$ 的样本，像以前那样，假定该随机样本是取自一个未知均值 μ 和方差 σ^2 的正态分布. 要检验以下假设

$$H_0:\mu\leqslant 1.2,$$
$$H_1:\mu>1.2.$$

   a. 在水平 $\alpha_0=0.05$ 的条件下检验假设.

   b. 计算 p-值.

2. 假定随机选择 9 个观察值，它们都来自于均值 μ 和方差 σ^2 都未知的正态分布. 已知：$\overline{X}_n=22$，$\sum_{i=1}^{n}(X_i-\overline{X}_n)^2=72$.

   a. 显著性水平取为 0.05，检验假设

$$H_0:\mu\leqslant 20,$$
$$H_1:\mu>20.$$

   b. 显著性水平取为 0.05，用双边 t-检验来检验假设

$$H_0:\mu=20,$$
$$H_1:\mu\neq 20.$$

   c. 根据数据，给出 μ 的置信系数为 95% 的置信区间.

3. 某种汽车的制造商声称在一般的城市驾驶条件下，汽车平均以不低于每加仑汽油 20 英里的水平行驶. 这种汽车的一个拥有者记录下在她城市驾驶的情况下的每加仑汽油行驶里程数的一些数据. 这些数据是在 9 种不同的场合为油箱加油时获得的. 每加仑汽油行驶的英里数的数据结果如下：

$$15.6,\ 18.6,\ 18.3,\ 20.1,\ 21.5,\ 18.4,\ 19.1,\ 20.4,\ 19.0$$

   给出在 $\alpha_0=0.05$ 下的假设检验，以此来检验厂商的声明. 仔细给出你必须要设的前提条件.

4. 设有 8 个观测值的随机样本，它取自服从均值 μ 和方差 σ^2 都未知的正态分布. 检验假设：

$$H_0:\mu=0,$$
$$H_1:\mu\neq 0.$$

已知：$\sum_{i=1}^{8} X_i = -11.2$ 和 $\sum_{i=1}^{8} X_i^2 = 43.7$. 如果在显著性水平0.1下进行一个对称的 t 检验，使得每个临界区域的尾部有0.05的概率，问接受还是拒绝 H_0?

5. 接习题4，检验水平仍为0.1. 但此时的 t 检验不是对称的，且在 $U \leqslant c_1$ 或 $U > c_2$ 的情况下都要拒绝 H_0，这里 $P(U \leqslant c_1) = 0.01, P(U > c_2) = 0.09$. 对于习题4中的样本数据，是否接受 H_0?

6. 设 $X_1, \cdots, X_n$ 是来自于均值 μ 和方差 σ^2 都未知的正态分布的随机样本. 在给定显著性水平 α_0 条件下，做以下假设的 t 检验：

$$H_0: \mu \leqslant \mu_0,$$
$$H_1: \mu > \mu_0.$$

令 $\pi(\mu, \sigma^2 \mid \delta)$ 表示该 t 检验的功效函数. 假设 $(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)$ 是满足：

$$\frac{\mu_1 - \mu_0}{\sigma_1} = \frac{\mu_2 - \mu_0}{\sigma_2}.$$

的值. 证明：$\pi(\mu_1, \sigma_1^2 \mid \delta) = \pi(\mu_2, \sigma_2^2 \mid \delta)$.

7. 考虑均值 μ 和方差 σ^2 都未知的正态分布，检验假设：

$$H_0: \mu \leqslant \mu_0,$$
$$H_1: \mu > \mu_0.$$

设从该分布中只能得到 X 的一个观测值. 但是可以利用来自于另一个正态分布的一组独立随机样本 $Y_1, \cdots Y_n$，它们与 X 有相同的方差 σ^2，且均值为0. 说明如何用自由度为 n 的 t 分布来检验假设 H_0 与 H_1?

8. 设 $X_1, \cdots, X_n$ 是来自于正态分布的随机样本，其均值 μ 和方差 σ^2 都未知. 又设 σ_0^2 是一个给定的正值. 在指定的检验水平 $\alpha_0 (0 < \alpha_0 < 1)$ 下检验假设：

496

$$H_0: \sigma^2 \leqslant \sigma_0^2,$$
$$H_1: \sigma^2 > \sigma_0^2.$$

令 $S_n^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$，并指明在假设检验过程中当 $S_n^2 / \sigma_0^2 \geqslant c$ 时，应当拒绝 H_0. 另外，设 $\pi(\mu, \sigma^2 \mid \delta)$ 代表检验过程的功效函数. 解释如何选择常数 c，使得无论 μ 取何值，下述要求总能满足：当 $\sigma^2 < \sigma_0^2$ 时，$\pi(\mu, \sigma^2 \mid \delta) < \alpha_0$；当 $\sigma^2 = \sigma_0^2$ 时，$\pi(\mu, \sigma^2 \mid \delta) = \alpha_0$；当 $\sigma^2 > \sigma_0^2$ 时，$\pi(\mu, \sigma^2 \mid \delta) > \alpha_0$.

9. 设10个观测值 $X_1, \cdots, X_n$ 来自于均值 μ 和方差 σ^2 都未知的正态分布的随机样本. 检验假设：

$$H_0: \sigma^2 \leqslant 4,$$
$$H_1: \sigma^2 > 4.$$

假定以第8题的形式进行检验. 当 $\alpha_0 = 0.05, S_n^2 = 60$ 时，是否接受 H_0?

10. 假定同习题9，从均值 μ 和方差 σ^2 都未知的正态分布中抽取10个观测值的随机样本，在显著性水平为0.05的条件下检验假设：

$$H_0: \sigma^2 = 4,$$
$$H_1: \sigma^2 \neq 4.$$

设当 $S_n^2 \leqslant c_1$ 或 $S_n^2 \geqslant c_2$ 时，拒绝 H_0. 此处 c_1, c_2 满足：当 H_0 为真时，

$$\Pr(S_n^2 \leqslant c_1) = \Pr(S_n^2 \geqslant c_2) = 0.025.$$

试求 c_1、c_2 的值.

11. 假设 U_1 有一个自由度为 m、非中心参数为 ψ 的非中心 t 分布. U_2 有一个自由度为 m、非中心参数为 $-\psi$ 的非中心 t 分布. 证明：$P(U_1 \geqslant c) = P(U_2 \leqslant -c)$.
12. 设 $X_1, \cdots, X_n$ 是来自于均值 μ 和方差 σ^2 都未知的正态分布的随机样本. 检验假设：
$$H_0: \mu \leqslant 3,$$
$$H_1: \mu > 3.$$
设样本大小 $n=17$，由样本观测值得到 $\overline{X}_n = 3.2$，$(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0.09$. 计算统计量 U 的值及相应的 p-值.
13. 考虑习题 12 的条件，若 $n=170$，同样由样本观测值得到 $\overline{X}_n = 3.2$，$(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0.09$. 计算统计量 U 的值及相应的 p-值.
14. 考虑习题 12 的条件，若要检验假设：
$$H_0: \mu = 3.1,$$
$$H_1: \mu \neq 3.1.$$
设样本大小 $n=17$，同样由样本观测值得到 $\overline{X}_n = 3.2$，$(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0.09$. 计算统计量 U 的值及相应的 p-值.
15. 考虑习题 14 的条件，现在设样本大小为 $n=170$，同样由样本观测值得到 $\overline{X}_n = 3.2$，$(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0.09$. 计算统计量 U 的值及相应的 p-值.
16. 考虑习题 14 的条件，现在设样本大小 $n=17$，同样由样本观测值得到 $\overline{X}_n = 3.2$，$(1/n)\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0.09$. 计算统计量 U 的值及相应的 p-值.
17. 证明：对于假设(8.3.6)的似然比检验是这样一个双边的 t 检验：在 $|U| \geqslant c$ 时拒绝 H_0，此处 U 由式(8.1.18)规定. 该问题的讨论比书上的单边检验的情形稍微简单一点，但又极其类似.
18. 证明：对假设(8.3.5)的似然比检验，当 $\mu \leqslant c$ 时，拒绝 H_0，其中 U 用等式(8.1.18)定义.
19. 设 Λ' 由式(8.3.17)定义. 考察似然比检验的另一种形式，该检验在 $\Lambda'(\boldsymbol{x}) \leqslant k'$ 时拒绝原假设 H_0.
    a. 令 $k' < 1$ 是一个常数，证明：$\Lambda'(\boldsymbol{x}) \leqslant k'$ 当且仅当 $\Lambda'(\boldsymbol{x}) \geqslant 1/k'$；
    b. 若 $k' \geqslant 1$，证明：对于所有数据 x，$\Lambda'(\boldsymbol{x}) \leqslant k'$. 这也就是为什么似然比检验的替换形式从来不用 $k' \geqslant 1$.

497

8.4 两个正态分布均值的比较

我们常常要比较两个分布，看哪个的均值较大或这两个均值是否有大的差异. 当两个分布都是正态分布时，基于 t 分布的检验和置信区间与考虑单个分布时的情形非常相似.

8.4.1　两个样本的 t 检验

现在考虑一个问题，随机样本是来自于两个方差未知但相等的正态分布，我们想知道哪个分布的均值较大. 特别地，设来自于一个正态分布的随机变量 $X_1,\cdots,X_n$ 形成一个有 m 个观察值的随机样本，它的均值 μ_1 和方差 σ^2 都是未知的；来自于另一个正态分布的随机变量 $Y_1,\cdots,Y_n$ 形成一个有 n 个观察值的独立随机样本，它的均值 μ_2 和方差 σ^2 也都是未知的. 虽然方差 σ^2 的值是未知的，但假设两个分布的方差 σ^2 是相同的. 如果最后这个假设得不到保证的话，那么该双样本的 t 检验是不合适的. 对于两个总体可能有不同方差的情形，在本节的后面部分将讨论一种不同的检验过程. 在8.5节中，将讨论一些用于比较两个正态分布的方差的过程，其中包括检验方差相等的假设.

假定在显著性水平 $\alpha_0(0<\alpha_0<1)$ 的情形下检验下面的假设

$$H_0:\mu_1\leqslant\mu_2,\qquad H_1:\mu_1>\mu_2. \tag{8.4.1}$$

对于每个检验步骤 δ，令 $\pi(\mu_1,\mu_2\sigma^2\mid\delta)$ 表示功效函数. 在本节的最后，我们导出似然比值检验. 直观地，如果 $\overline{X}_m-\overline{Y}_n$ 较大，则拒绝 H_0 的假设是合理的. 根据8.3节的 t 检验，定义

$$S_X^2=\sum_{i=1}^m(X_i-\overline{X}_m)^2 \text{ 和 } S_Y^2=\sum_{i=1}^n(Y_i-\overline{Y}_n)^2. \tag{8.4.2}$$

于是，使用的检验统计量为：

$$U=\frac{(m+n-2)^{1/2}(\overline{X}_m-\overline{Y}_n)}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}(S_X^2+S_Y^2)^{1/2}}. \tag{8.4.3}$$

对于某常数 c，如果 $U\geqslant c$，则双样本 t 检验将拒绝 H_0. 下面，我们将给出 U 的分布.

8.4.2　两样本 t 检验的性质

498

对于每对 μ_1、μ_2 的值和每个 σ^2，样本均值 $\overline{X}_m$ 服从均值为 μ_1、方差为 σ^2/m 的正态分布，且样本均值 $\overline{Y}_n$ 服从均值为 μ_2、方差为 σ^2/n 的正态分布. 由于 $\overline{X}_m$ 和 $\overline{Y}_n$ 相互独立，差 $\overline{X}_m-\overline{Y}_n$ 服从均值为 $\mu_1-\mu_2$、方差为 $[(1/m)+(1/n)]\sigma^2$ 的正态分布. 因此，当 $\mu_1=\mu_2$ 时，下面的随机变量 Z 服从标准正态分布：

$$Z=\frac{\overline{X}_m-\overline{Y}_n}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}\sigma}. \tag{8.4.4}$$

而且，对于所有的 μ_1，μ_2 和 σ^2 的值，随机变量 S_X^2/σ^2 服从自由度为 $m-1$ 的 χ^2 分布：随机变量 S_Y^2/σ^2 服从自由度为 $n-1$ 的 χ^2 分布，并且这两个随机变量相互独立. 因此，下面的随机变量 W 服从自由度为 $m+n-2$ 的 χ^2 分布：

$$W = \frac{S_X^2 + S_Y^2}{\sigma^2}. \tag{8.4.5}$$

此外，$\overline{X}_m$、$\overline{Y}_n$、S_X^2 和 S_Y^2 这四个变量相互独立. 该结论从以下的两个事实中得到：（i）如果一个随机向量只是 $X_1,\cdots,X_n$ 的一个函数，而另一个随机向量只是 $Y_1,\cdots,Y_n$的函数，那么这两个向量必须独立.（ii）由定理7.3.1，$\overline{X}_m$ 与 S_X^2 独立，$\overline{Y}_n$ 与 S_Y^2 也独立. 从而随机变量 Z 与 W 独立. 当$\mu_1=\mu_2$ 时，随机变量 Z 服从标准正态分布，对于所有的μ_1、μ_2 和 σ^2 值，随机变量 W 服从自由度为 $m+n-2$ 的χ^2 分布，下面可以把统计量 U 表示成以下的形式：

$$U = \frac{Z}{[W/(m+n-2)]^{1/2}}. \tag{8.4.6}$$

因此，当$\mu_1=\mu_2$ 时，根据7.4节中 t 分布的定义，可得 U 服从自由度为 $m+n-2$ 的 t 分布. 于是，当$\mu_1=\mu_2$ 时，如果 c 是自由度为 $m+n-2$ 的 t 分布的分位数 $1-\alpha_0$，即 $c=T_{m+n-2}^{-1}(1-\alpha_0)$，那么 $\Pr(U\geqslant c)=\alpha_0$. 不论$\mu_1$ 和μ_2 的共同值以及 σ^2 的值是什么，这个结论总是正确的.

如果$\mu_1<\mu_2$，那么像在8.3节中讨论的一样，可以证明：无论 σ^2 取什么值，都有 $\Pr(U\geqslant c)<\alpha_0$. 于是当 $U>c$ 时，拒绝 H_0 的检验过程在 $U>c$ 时有检验水平 α_0.

注： 另一个单边假设. 如果原假设是 $H_0:\mu_1\geqslant\mu_2$，在相应的水平 α_0 下 t 检验将在 $U\leqslant -T_{m+n-2}^{-1}(1-\alpha_0)$时拒绝原假设. 该检验具有与另一单边检验相类似的性质.

例8.4.1 在英国的罗马陶器 Tubb，Parker 和 Nickless（1980）阐述了一项研究，该项研究是关于在英国各地区被发现的产自罗马时代的陶器样本，对于每个陶器样本都进行的一项测量是测量样本中氧化铝的百分含量. 假设我们感兴趣的是比较两个不同的地点的样本中氧化铝的百分含量. 对来自于样本容量为 $m=14$ 的 Llanederyn 的样本进行分析得到样本平均值 $\overline{X}_m=12.56$ 和方差 $S_X^2=24.65$. 对另一个来自于样本容量为 $n=5$ 的 Ashley Rails 样本的分析得到 $\overline{Y}_n=17.32$ 和 $S_Y^2=11.01$. 假设将数据用正态随机变量建模，它具有两个不同的均值μ_1 和μ_2，但是方差都是 σ^2. 希望检验原假设 $H_0:\mu_1\geqslant\mu_2$ 和备择假设 $H_1:\mu_1<\mu_2$. 等式(8.4.3)定义的 U 的观测值为 -6.302. 从书中 t 分布表中查看自由度为 $m+n-2=17$ 的分布，发现 $T_{17}^{-1}(0.995)=2.898$ 和 $U<-2.898$. 所以将在任何显著性水平 $\alpha_0\geqslant 0.005$ 的情况下拒绝原假设 H_0. 实际上，与此 U 值相应的 p 值大约是 4×10^{-6}. ◀

499

8.4.3 检验的功效

双样本 t 检验的功效函数具有以下五个性质：

i. $\pi(\mu_1,\mu_2,\sigma^2|\delta)=\alpha_0$，当$\mu_1=\mu_2$ 时，ii. $\pi(\mu_1,\mu_2,\sigma^2|\delta)<\alpha_0$，当$\mu_1<\mu_2$ 时，

iii. $\pi(\mu_1,\mu_2,\sigma^2|\delta)>\alpha_0$，当$\mu_1>\mu_2$ 时，iv. $\pi(\mu_1,\mu_2,\sigma^2|\delta)\to0$，当$\mu_1-\mu_2\to-\infty$ 时，

v. $\pi(\mu_1,\mu_2,\sigma^2|\delta)\to1$，当$\mu_1-\mu_2\to\infty$ 时.

与单样本的情况(见8.3节)相类似，可以证明：无论 σ^2 取什么值，如果 $\mu_1 > \mu_2$，那么 $\Pr(U \geq c) > \alpha_0$. 此外，需重申的是，上面最后两个性质较难证明.

对于每个 $\mu_1 - \mu_2$ 的值，两样本 t 检验的功效函数可以使用在8.3节中介绍过的非中心的 t 分布来计算. 与8.3节的推理几乎完全相同，可以证明等式(8.4.6)中的 U 的分布是一个自由度为 $m+n-2$，非中心参数为

$$\psi = \frac{\mu_1 - \mu_2}{\sigma\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2}}. \tag{8.4.7}$$

的非中心 t 分布.

如果我们手头没有合适的计算机程序，可以用图8-7来近似功效函数的计算. 如果Llanederyn的均值比Ashley Rails的均值小 1.5σ，那么 $\psi = 1.5/(1/14 + 1/5)^{1/2} = 2.88$. 在图8-7的右边我们可以看到原假设 $H_0: \mu_1 \geq \mu_2$ 的0.01水平检验的功效大约是0.65(实际功效是0.63).

8.4.4　双边备择

可以容易地将双样本 t 检验修正为在一个特定的显著性水平 α_0 下检验假设

500

$$\begin{aligned} H_0&: \mu_1 = \mu_2, \\ H_1&: \mu_1 \neq \mu_2. \end{aligned} \tag{8.4.8}$$

如果 $|U| \geq c$，其中 $c = T^{-1}_{m+n-2}(1 - \alpha_0/2)$，$\alpha_0$ 水平双边 t 检验拒绝原假设 H_0. 该检验是建立在式(8.4.3)中定义的统计量 U 的基础之上的. 当 $U = u$ 时，p-值等于 $2[1 - T_{m+n-2}(|u|)]$(参见习题9).

例8.4.2　铜矿石的比较　假定一个八种矿石的随机样本是来自于某一个地方的铜矿，并且用克来做测量单位表示每一种矿石中铜的含量. 用 $X_1, \cdots, X_8$ 来表示这8个量，并假设它们有如下观察结果：$\overline{X}_8 = 2.6$ 和 $S_X^2 = 0.32$. 同时也假设第二个随机样本由10种矿石样本组成，来自于矿址的另一个部分. 我们用 $Y_1, \cdots, Y_{10}$ 来表示这10个量，并假定观察结果为 $\overline{Y}_{10} = 2.3$ 和 $S_Y^2 = 0.22$. 用 μ_1 表示来自于第一个地点的矿石的铜含量的数据平均值. 用 μ_2 表示来自于第二个地点的矿石的铜含量的数据的平均值. 下面检验假设(8.4.8).

假定所有的观测值都是服从正态分布的，且两个地方相应的方差是相同的，即使均值可能不同. 在本例中样本容量是 $m=8$ 和 $n=10$，且在(8.4.3)中定义的统计量 U 的值是3.442. 另外，查看自由度为16的 t 分布，发现 $T_{16}^{-1}(0.995) = 2.921$，所以相应于该 U 观察值的尾部区域小于 2×0.005. 因此，在任何一个显著性水平 $\alpha_0 \geq 0.01$ 下都要拒绝原假设（事实上当 $U = 3.442$，双边尾部区域是0.003). ——◀

双边双样本 t 检验的功效函数和单样本双边 t 检验的功效函数一样，是建立在非中心 t 分布的基础之上的. 这种在 $|U| \geq c$ 时拒绝 $H_0: \mu_1 = \mu_2$ 的检验 δ 满足：

$$\pi(\mu_1,\mu_2,\sigma^2|\delta)=T_{m+n-2}(-c|\psi)+1-T_{m+n-2}(c|\psi),$$

其中 $T_{m+n-2}(\cdot|\psi)$ 是自由度为 $m+n-2$、非中心参数由式(8.4.7)给出的非中心 t 分布的密度函数. 图8-9可以用来近似这个功效函数.

8.4.5 扩展研究：不相等方差

t 检验可以推广到用于两组服从正态分布的数据方差不相同、但方差之比已知的情形. 具体来说，假定 $X_1,\cdots,X_m$ 构成一组随机样本，来自于服从均值为 μ_1、方差为 σ_1^2 的正态分布，$Y_1,\cdots,Y_n$ 为另一组独立的随机样本，来自于均值为 μ_2、方差为 σ_2^2 的正态分布，若 μ_1、μ_2、σ_1^2 和 σ_2^2 均未知，但已知 $\sigma_2^2=k\sigma_1^2$, k 为一已知的正常数，那么可以证明，当 $\mu_1=\mu_2$ 时(见本节最后的习题4)，以下的随机变量 U 服从自由度为 $m+n-2$ 的 t 分布 [501]

$$U=\frac{(m+n-2)^{1/2}(\overline{X}_m-\overline{Y}_n)}{\left(\frac{1}{m}+\frac{k}{n}\right)^{1/2}\left(S_X^2+\frac{S_Y^2}{k}\right)^{1/2}}.\tag{8.4.9}$$

因此，U 可以将(8.4.9)定义的统计量用于检验(8.4.1)的假设或(8.4.8)的假设.

最后，如果所有4个变量 μ_1、μ_2 和 σ_1^2、σ_2^2 的值都是未知的，且 σ_1^2 和 σ_2^2 的比值也未知，那么对于(8.4.1)和(8.4.8)两式的假设检验的问题会变得很困难. 甚至似然比统计量 Λ 没有已知的分布. 我们称这个问题之为贝伦斯－费歇尔问题. 对于贝伦斯－费歇尔问题的一些模拟解法会在第11章(例11.1.4和例11.4.9)给出. 人们还提出了一些其他的检验方法，但它们大多数都遭遇到合理性和可用性的争论. 在这些方法中，最为人熟知的是威尔齐(1938, 1947, 1951)发表的一系列文章给出的. 威尔齐提出使用统计量

$$V=\frac{\overline{X}_m-\overline{Y}_n}{\left(\frac{S_X^2}{m(m-1)}+\frac{S_Y^2}{n(n-1)}\right)^{1/2}}.\tag{8.4.10}$$

即使在 $\mu_1=\mu_2$ 时，V 分布的形式也是未知的，但是，威尔齐却运用了以下 t 分布的方法来近似 V 分布. 令

$$W=\frac{S_X^2}{m(m-1)}+\frac{S_Y^2}{n(n-1)},\tag{8.4.11}$$

并用与 W 有相同均值和方差的伽玛分布来近似表示 W 分布（见习题12). 如果我们现在假定 W 的确服从该伽玛分布，则变量 V 就有自由度为

$$\frac{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{\sigma_1^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{\sigma_2^2}{n}\right)^2}.\tag{8.4.12}$$

的 t 分布. 接下来，用无偏估计量 $s_x^2/(m-1)$ 和 $s_y^2/(n-1)$ 分别取代 σ_1^2 和 σ_2^2. 代入到等式(8.4.12)中，则可以得到所需威尔齐 t 分布近似的自由度

$$v=\frac{\left(\frac{s_x^2}{m(m-1)}+\frac{s_y^2}{n(n-1)}\right)^2}{\frac{1}{(m-1)^3}\left(\frac{s_x^2}{m}\right)^2+\frac{1}{(n-1)^3}\left(\frac{s_y^2}{n}\right)^2}. \tag{8.4.13}$$

502

在等式(8.4.13)中，s_x^2 和 s_y^2 是 S_X^2 和 S_Y^2 的观察值，为了概括威尔齐的过程，假想式(8.4.10)中的变量 V 在 $\mu_1=\mu_2$ 时服从自由度为 v 的 t 分布，通过把 V 和自由度为 v 的 t 分布的各种分位数比较来构造出单边和双边的假设检验. 如果 v 不是一个整数，将它近似为离得最近的整数，或者使用计算机程序来处理非整数自由度的 t 分布.

例 8.4.3　比较铜矿石　使用例 8.4.2 中的数据，通过计算可得

$$V=\frac{2.6-2.3}{\left(\frac{0.32}{8\times 7}+\frac{0.22}{10\times 9}\right)^{1/2}}=3.321,$$

$$v=\frac{\left(\frac{0.32}{8\times 7}+\frac{0.22}{10\times 9}\right)^2}{\frac{1}{7^3}\left(\frac{0.32}{8}\right)^2+\frac{1}{9^3}\left(\frac{0.22}{10}\right)^2}=12.49.$$

与观察数据相对应的 p-值为 $2[1-T_{12.49}(3.321)]=0.0058$，与在 8.4.2 中所得的数据没有太大差别. ◀

8.4.6　扩展研究：作为似然比检验的两样本的 t 检验

在本节中，将会看到对于假设(8.4.1)的两样本的 t 检验是一个似然比检验. 在两个样本中的 $x_1,\cdots,x_m$ 和 $y_1,\cdots,y_n$ 观察值取定后，似然函数 $g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma^2)$ 是：

$$g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma^2)=f_m(\boldsymbol{x}|\mu_1,\sigma^2)f_n(\boldsymbol{y}|\mu_2,\sigma^2).$$

这里，$f_m(\boldsymbol{x}|\mu_1,\sigma_2)$ 和 $f_n(\boldsymbol{y}|\mu_2,\sigma^2)$ 具有(8.3.10)给出的形式，且在两项中 σ_2 的值是相同的，在该种情况下，$\Omega_0=\{(\mu_1,\mu_2,\sigma^2)|\mu_1\leqslant\mu_2\}$ 和 $\Omega_1=\{(\mu_1,\mu_2,\sigma^2)|\mu_1>\mu_2\}$，则似然比统计量为：

$$\Lambda(\boldsymbol{x},\boldsymbol{y})=\frac{\sup_{\{(\mu_1,\mu_2,\sigma^2)|\mu_1>\mu_2\}}g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma^2)}{\sup_{\{(\mu_1,\mu_2,\sigma^2)|\mu_1\leqslant\mu_2\}}g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma^2)}. \tag{8.4.14}$$

k 的取值一般要满足使得检验有所希望的显著水平 α_0 的要求. 在 $\Lambda(\boldsymbol{x},\boldsymbol{y})\geqslant k$ 的情况下，则似然比检验就认定拒绝原假设 H_0.

为了能方便地在(8.4.14)中取得最大值，令

$$s_x^2=\sum_{i=1}^{m}(x_i-\bar{x}_m)^2 \text{ 和 } s_y^2=\sum_{i=1}^{n}(y_i-\bar{y}_n)^2.$$

由此可将原式写为：

$$g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma^2)$$

$$=\frac{1}{(2\pi\sigma^2)^{(m+n)/2}}\exp\left(-\frac{1}{2\sigma^2}[m(\bar{x}_m-\mu_1)^2+n(\bar{y}_n-\mu_2)^2+s_x^2+s_y^2]\right).$$

503

当 $\bar{x}_m \geqslant \bar{y}_n$ 时，容易看出(8.4.14)的分子取最大值发生在

$$\mu_1=\bar{x}_m,\mu_2=\bar{y}_n \text{ 和 } \sigma^2=\frac{1}{m+n}(s_x^2+s_y^2). \tag{8.4.15}$$

(8.4.15)中的各个值当 $\bar{x}_m<\bar{y}_n$ 时也能使分母取得最大值. 对于另外两种情况(分子在 $\bar{x}_m<\bar{y}_n$ 时和分母在 $\bar{x}_m\geqslant\bar{y}_n$ 时)不难发现为了得到最大值，$\mu_1=\mu_2$ 是必需的. 在该种情况中，当

$$\mu_1=\mu_2=\frac{\overline{m x}_m+\overline{n y}_n}{m+n},$$

$$\sigma^2=\frac{mn(\bar{x}_m-\bar{y}_n)^2/(m+n)+s_x^2+s_y^2}{m+n}.$$

时，可以达到最大值，将上面的值代入(8.4.14)的式中，就可以得到：

$$\Lambda(x,y)=\begin{cases}(1+v^2)^{(m+n)/2} & \text{若 } \bar{x}_m\geqslant\bar{y}_n,\\ (1+v^2)^{-(m+n)/2} & \text{若 } \bar{x}_m<\bar{y}_n.\end{cases}$$

这里

$$v=\frac{(\bar{x}_m-\bar{y}_n)}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}(s_x^2+s_y^2)^{1/2}}. \tag{8.4.16}$$

由此可以直接证明 $\Lambda(x,y)\geqslant k$ 等价于 $v\geqslant k'$，其中 k' 是另外某个常数，最后，注意到 $(m+n-2)^{1/2}v$ 是 U 的观察值，所以当进行似然比检验时，满足 $U\geqslant c$ 时拒绝原假设，c 为某个常数. 而这与两样本 t 检验的结论是相同的. 很容易地能将前面所做的讨论修改后用到处理其他单边检验和双边检验的情形(双边检验的情形见习题 13).

8.4.7 小结

假设观察来自两个正态分布的独立的随机样本：$X_1,\cdots,X_m$ 的均值为 μ_1，方差为 σ_1^2，而 $Y_1,\cdots,Y_n$ 的均值为 μ_2，方差为 σ_2^2，为了对 μ_1 和 μ_2 进行假设检验，如果假定 $\sigma_1^2=\sigma_2^2$，那么就可以用 t 检验. 所有的 t 检验都是利用式(8.4.3)中所定义的统计量 U. 在水平 α_0 下检验 $H_0:\mu_1=\mu_2$ 和 $H_1:\mu_1\neq\mu_2$ 时，如果 $|U|\geqslant T_{m+n-2}^{-1}(1-\alpha_0/2)$，则拒绝 H_0. 其中 $T_{m+n-2}^{-1}(1-\alpha_0/2)$ 是自由度为 $m+n-2$ 的 t 分布中的分位数函数. 在水平 α_0 下检验 $H_0:\mu_1\leqslant\mu_2$ 和 $H_1:\mu_1>\mu_2$ 时，若 $U>T_{m+n-2}^{-1}(1-\alpha_0)$ 则拒绝 H_0. 在水平 α_0 下检验 $H_0:\mu_1\geqslant\mu_2$ 和 $H_1:\mu_1<\mu_2$ 时，若 $U<-T_{m+n-2}^{-1}(1-\alpha_0)$ 则拒绝 H_0. 这些检验中的功效函数可用非中心的 t 分布来计算，如果不假定 $\sigma_1^2=\sigma_2^2$，可以用近似检验.

504

习题

1. 我们要比较人工降雨云和非人工降雨云的降水量的对数(ln－降水量)的均值，正如例7.3.1所讨论过的. 令$X_1,\cdots,X_{26}$为26朵人工降雨云的ln－降水量，再令$Y_1,\cdots,Y_{26}$为26朵非人工降雨云的ln－降水量. 样本统计量的观察值为$\overline{X}_{26}=5.134$，$\overline{Y}_{26}=3.990$，$S_X^2=63.96$和$S_Y^2=67.39$. 原假设是人工降雨云所产生的降水并不比非人工降雨云多，而备择假设与此相反. 在显著性水平为$\alpha_0=0.01$下检验这些假设. 假定所有ln－降水量的方差都是一样的.
2. 假设把第一种药A用于随机选出的8个病人，经过了一段固定时间以后对病人身体里细胞中药的浓度按适当的单位进行测量，假设8个病人所测得的浓度如下所示：

   1.23，1.42，1.41，1.62，1.55，1.51，1.51，1.60和1.76

   假设还有第二种药B用于随机选出的6个不同的病人，然后用同样的方法从这6个人身上测量药B的浓度，结果如下：

   1.76，1.41，1.87，1.49，1.67和1.81

   假定这些观测资料服从有同样未知方差的正态分布，在显著性0.10的水平下检验下面的假设. 原假设是所有病人中药A的浓度平均值至少与药B的浓度平均值一样大，备择假设是药B的浓度平均值比药A的大.
3. 重新考虑习题2的情况，但现在变为要求检验以下假设：原假设为在所有的病人中药A浓度平均值和药B的相等. 双侧的备择假设是两种药的浓度平均值不一样. 找到数c使得在水平0.05的双侧t检验中，当$|U|\geq c$时拒绝H_0，其中U由等式(8.4.3)定义. 并且进行该检验.
4. 假设一个随机样本$X_1,\cdots,X_m$来自于一个均值为μ_1和方差为σ_1^2的正态分布，独立的随机样本$Y_1,\cdots,Y_n$来自于另一个均值为μ_2和方差为σ_2^2的正态分布. 证明：当$\mu_1=\mu_2$且$\sigma_2^2=k\sigma_1^2$时，式(8.4.9)中所定义的随机变量U服从自由度为$m+n-2$的t分布.
5. 重新考虑习题2的条件和观察值. 然而，现在假设药A的每个观察值有一个未知方差σ_1^2，药B的每个观察值也有一个未知方差σ_2^2，但已知$\sigma_2^2=(6/5)\sigma_1^2$，在显著性水平0.10的情况下检验习题2中所给出的假设.
6. 假设一个随机样本$X_1,\cdots,X_m$来自于均值μ_1未知和方差σ^2未知的正态分布，独立的随机样本$Y_1,\cdots,Y_n$来自于另一个未知均值μ_2和同样的未知方差σ^2的正态分布. 对于每一个常数$\lambda(-\infty<\lambda<\infty)$，构造一个自由度为$m+n-2$的$t$检验来检验以下假设：

$$H_0:\mu_1-\mu_2=\lambda,$$
$$H_1:\mu_1-\mu_2\neq\lambda.$$

7. 重新考虑习题2的条件，令μ_1为药A的每次观察的均值，令μ_2表示药B的每次观察的均值，像在习题2中所做的那样，我们假设所有的观察值有相同但未知的方差. 利用习题6的结果为$\mu_1-\mu_2$构造一个置信系数为0.90的置信区间.
8. 在例8.4.2中确定当$|\mu_1-\mu_2|=\sigma$时，水平为0.01的检验的功效.
9. 假定我们希望检验假设(8.4.8)，我们要用(8.4.3)中所确定的统计量U，且当$|U|$值大时，拒绝H_0. 证明：当观察到$U=u$时，p-值为$2[1-T_{m+n-2}(|u|)]$.

10. Lyle 等人(1987)进行了一项试验来研究补钙对于非洲裔美国男人血压的影响. 有一组 10 个人得以补充钙，而另一个组的 11 人得到的是安慰剂，试验持续了 12 周，在这 12 周之前和之后，都测量了每个对象在休息状态下的收缩压. 表 8-1 给出的是 12 周后减去 12 周前所得的差. 原假设为：补钙组的血压平均变化要比安慰剂组的血压平均变化低. 试用水平 $\alpha=0.1$ 来检验这个假设.

505

表 8-1 习题 10 中的血压资料

钙	7	−4	18	17	−3	−5	1	10	11	−2	
安慰剂	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

11. Frisby 和 Clatworthy(1975)研究了人破解随机点立体图识别物体所需要的时间. 随机点立体图指的是成对的图像，它们一开始出现的时候好像是些随机点. 实验的对象从一个适当的距离看成对的图像，眼睛恰好跨越一正确的距离，于是两个像就合成为一个可以辨认的物体. 实验关心的是在多大程度上，事先关于那个可以辨认的物体的信息会影响两个图像的合成时间. 一个组的 43 位实验对象在合成图像之前没有看过该物体的图片. 他们的平均时间 $\overline{X}_{43}=8.560$ 而 $S_X^2=2745.7$. 第二个组有 35 位实验对象，他们事先看过图片，结果发现他们的样本统计量 $\overline{Y}_{35}=5.551$ 和 $S_Y^2=783.9$. 原假设是第一组的平均时间不比第二组的时间长，而备择假设为第一组的平均时间长一些
    a. 在显著性水平 $\alpha_0=0.01$ 下检验假设，假定两组的方差相等；
    b. 在显著性水平 $\alpha_0=0.01$ 下检验假设，用威尔士近似检验.
12. 求式(8.4.11)中随机变量 W 的均值 a 和方差 b. 现在，令 a 和 b 分别是参数为 α 和 β 的伽玛分布的均值和方差，证明：2α 和(8.4.12)的表达式一样.
13. 令 U 为式(8.4.3)所定义的形式，假定要求对(8.4.8)中的假设进行检验. 证明：每个似然比检验有这样的形式，即若 $|U|\geq c$ 时，拒绝 H_0，其中 c 为常数. 提示：首先证明 $\Lambda(\boldsymbol{x},\boldsymbol{y})=(1+v^2)^{-(m+n)/2}$，$v$ 由式(8.4.16)所确定.

8.5 F 分布

在本节中，我们介绍 F 分布族. 该族可用于两种不同情况的假设检验. 第一种情况是当我们想要检验关于两种不同正态分布的方差的假设. 本节将导出这些检验，它们基于有 F 分布的统计量. 第二种情况将在第十章讨论检验有关多于两个的正态分布的平均数的假设时进行.

8.5.1 F 分布的定义

在本节中我们要介绍一个概率分布，叫做 F 分布，它用在许多检验假设的重要问题中，对于两种或两种以上的正态分布，将从每个分布中抽取的随机样本进行比较. 我们从 F 分布定义和导出它的概率函数开始.

考虑两个独立的随机变量 Y 和 W，Y 是自由度为 m 的 χ^2 分布，W 服从自由度为 n 的 χ^2 分布，m 和 n 是正整数. 定义一个新的随机变量 X 如下：

506

$$X = \frac{Y/m}{W/n} = \frac{nY}{mW}. \tag{8.5.1}$$

称 X 的分布为自由度为 m 和 n 的 F 分布.

我们将证明如果一个随机变量 X 服从自由度为 m 和 n 的 F 分布，那么它的密度函数 $f(x)$ 如下，当 $x>0$ 时，

$$f(x) = \frac{\Gamma\left[\frac{1}{2}(m+n)\right]m^{m/2}n^{n/2}}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \cdot \frac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}}. \tag{8.5.2}$$

当然，当 $x \leqslant 0$ 时，$f(x)=0$.

既然随机变量 Y 和 W 是独立的，它们的联合密度函数 $g(y,w)$ 将是它们各自密度函数的乘积. 而且，因为 Y 和 W 都服从 χ^2 分布，由 χ^2 分布的概率密度函数(见 7.2 节)，可知 $g(y,w)$ 有如下形式，当 $y>0$ 且 $w>0$ 时，

$$g(y,w) = cy^{(m/2)-1}w^{(n/2)-1}e^{-(y+w)/2}, \tag{8.5.3}$$

其中

$$c = \frac{1}{2^{(m+n)/2}\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)}. \tag{8.5.4}$$

现在把变量 Y 和 W 换成 X 和 W，这里 X 是由(8.5.1)定义的. X 和 W 的联合密度函数 $h(x,w)$ 首先是在式(8.5.3)中用 x 和 w 的表达式代替 y，然后乘以 $|\partial y/\partial x|$ 得到的. 从式(8.5.1)得到 $y=(m/n)xw$ 和 $\partial y/\partial x=(m/n)w$. 因此，对于 $x>0$ 和 $w>0$，联合密度函数 $h(x,w)$ 有如下形式

$$h(x,w) = c\left(\frac{m}{n}\right)^{m/2} x^{(m/2)-1}w^{[(m+n)/2]-1}\exp\left[-\frac{1}{2}\left(\frac{m}{n}x+1\right)w\right]. \tag{8.5.5}$$

其中常数 c 就是式(8.5.4)中所给出的.

对于每个 $x>0$ 的值，X 的边际密度函数可以从关系

$$f(x) = \int_0^\infty h(x,w)\,dw. \tag{8.5.6}$$

中得到. 根据式(5.9.8)得

$$\int_0^\infty w^{[(m+n)/2]-1}\exp\left[-\frac{1}{2}\left(\frac{m}{n}x+1\right)w\right]dw = \frac{\Gamma\left[\frac{1}{2}(m+n)\right]}{\left[\frac{1}{2}\left(\frac{m}{n}x+1\right)\right]^{(m+n)/2}}. \tag{8.5.7}$$

根据式(8.5.4)和(8.5.7)得出结论，概率密度函数 $f(x)$ 具有式(8.5.2)给出的形式.

8.5.2 F 分布的性质

当讨论具有自由度 m 和 n 的 F 分布时，如同可以从式(8.5.1)中 X 的定义看到的那样，条件中的数字 m 和 n 的给出次序是很重要的. 当 $m \neq n$ 时，自由度 m 和 n 的 F 分布和自由度 n 和 m 的 F 分布是两个不同的分布. 事实上，如果一个随机变量 X 服从自由度 m 和 n 的 F 分布，那么它的倒数 $1/X$ 服从自由度 n 和 m 的 F 分布. 该陈述由式(8.5.1)中 X 被表示为两个随机变量的比值可得.

F 分布和 t 分布是有关的：如果一个随机变量 X 服从自由度为 n 的 t 分布，那么 X^2 就服从自由度是 1 和 n 的 F 分布. 该结论来自于式(7.4.1)中 X 的表示式.

在本书的最后给出了两个简短的 F 分布分位数表. 在表中只给出对于不同可能的一对 m 和 n 的值对应的0.95分位数和0.975分位数. 即，如果 G 指的是自由度 m 和 n 的 F 分布的分布函数，那么表中给出 x_1 和 x_2 的值是使 $G(x_1)=0.95$ 和 $G(x_2)=0.975$. 通过运用 X 的 F 分布和 $1/X$ 的 F 分布的关系，就有可能得到在0.05和0.025分位数的 F 分布.

例 8.5.1 确定在0.05分位数下的 F 分布 假如一个随机变量 X 服从自由度为6和12的 F 分布. 要确定当 $\Pr(X<x)=0.05$ 时 x 的值.

如果令 $Y=1/X$，那么 Y 将服从自由度为12和6的 F 分布. 在本书最后的表中可以找到 $\Pr(Y>4.00)=0.05$，由于关系 $Y>4.00$ 等价于 $X<0.25$，所以，$\Pr(X<0.25)=0.05$，因而 $x=0.25$. ◀

8.5.3 比较两个正态分布的方差

现在需要考虑使用 F 分布的假设检验的问题. 假设来自于一个正态分布的随机变量 $X_1,\cdots,X_m$ 形成一个由 m 个观察值所组成的随机样本，它的均值 μ_1 和方差 σ_1^2 都是未知的，并且来自于另一个正态分布的随机变量 $Y_1,\cdots,Y_n$ 形成一个有 n 个观察值的独立随机样本，它的均值 μ_2 和方差 σ_2^2 也都是未知的. 最后，假定要在置信水平 $\alpha_0(0<\alpha_0<1)$ 下检验下面的假设

$$\begin{aligned} H_0&: \sigma_1^2 \leqslant \sigma_2^2, \\ H_1&: \sigma_1^2 > \sigma_2^2. \end{aligned} \tag{8.5.8}$$

对于每个检验程序 δ，令 $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\,|\,\delta)$ 表示 δ 的功效函数，在本节的结尾，将得到似然比检验. 现在定义 S_X^2 和 S_Y^2 为式(8.4.2)所定义的平方和. 那么 $S_X^2/(m-1)$ 和 $S_Y^2/(n-1)$ 分别是 σ_1^2 和 σ_2^2 的估计量. 直观地知道如果 V 大，应拒绝 H_0，这里，[508]

$$V=\frac{S_X^2/(m-1)}{S_Y^2/(n-1)}. \tag{8.5.9}$$

称该种检验为 F 检验. 因为当 $\sigma_1^2=\sigma_2^2$ 时 V 的分布是一个 F 分布. 接下来会证明该事实.

8.5.4 F 检验的性质

由7.3节可知随机变量 S_X^2/σ_1^2 服从自由度为 $m-1$ 的 χ^2 分布，同时随机变量 S_Y^2/σ_2^2 服从自由度为 $n-1$ 的 χ^2 分布. 此外，由于他们是从两个独立的样本中计算出的，而这两个随机变量是互相独立的. 因此下面的随机变量 V^* 服从自由度为 $m-1$ 和 $n-1$ 的 F 分布：

$$V^* = \frac{S_X^2/[(m-1)\sigma_1^2]}{S_Y^2/[(n-1)\sigma_2^2]}. \tag{8.5.10}$$

由方程(8.5.9)和(8.5.10)可以看到 $V=(\sigma_1^2/\sigma_2^2)V^*$. 如果 $\sigma_1^2=\sigma_2^2$，则 $V=V^*$. 因此当 $\sigma_1^2=\sigma_2^2$ 时，统计量 V 将服从自由度为 $m-1$ 和 $n-1$ 的 F 分布. 在此种情况下可以通过使用 F 分布表来选择常数 c，使得 $\Pr(V\geqslant c)=\alpha_0$ 与 σ_1^2 和 σ_2^2 的共同值无关，也与 μ_1 与 μ_2 的值无关，事实上 c 是对应的 F 分布的 $1-\alpha_0$ 分位数.

现在设 $\sigma_1^2<\sigma_2^2$，则

$$\Pr(V\geqslant c)=\Pr\left(\frac{\sigma_1^2}{\sigma_2^2}V^*\geqslant c\right)=\Pr\left(V^*\geqslant\frac{\sigma_2^2 c}{\sigma_1^2}\right)<\Pr(V^*\geqslant c)=\alpha_0. \tag{8.5.11}$$

因此当 $\sigma_1^2<\sigma_2^2$ 时，$\Pr(V\geqslant c)<\alpha_0$ 与 μ_1 与 μ_2 的值无关，当 $V\geqslant c$ 时拒绝 H_0 的检验，其显著性水平为 α_0.

例8.5.2　进行 F 检验　从均值 μ_1 和方差 σ_1^2 未知的正态分布中选出6个观察值 $X_1,\cdots,X_6$，同时 $S_X^2=30$. 假设从均值 μ_2 和方差 σ_2^2 未知的正态分布中选出21个观察值 $Y_1,\cdots,Y_{21}$ 同时 $S_Y^2=40$. 对假设(8.5.8)进行 F 检验.

在本例中，$m=6$ 和 $n=21$，所以当 H_0 为真时，由式(8.5.9)所定义的统计量 V 服从自由度为5和20的 F 分布. 由式(8.5.9)看出，所给样本 V 的值

509

$$V=\frac{30/5}{40/20}=3.$$

从本书最后给出的表可以知道在分位数为0.95、自由度为5和20的 F 分布的值为2.71，在分位数为0.975时 F 的值是3.29. 因此，$V=3$ 所对应的尾部的值是小于0.05、大于0.025的，因此在 $\alpha_0=0.05$ 的置信水平上的 F 值拒绝 H_0: $\sigma_1^2\leqslant\sigma_2^2$. 在 $\alpha_0=0.025$ 的置信水平上的 F 值接受 H_0（使用计算机程序去计算 F 的密度函数所给出的 p-值为0.035）. ◀

8.5.5 检验的功效

刚才导出的 F 检验的功效函数有以下5种性质：

i. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|\delta)=\alpha_0$，当 $\sigma_1^2=\sigma_2^2$.

ii. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2|\delta)<\alpha_0$，当 $\sigma_1^2<\sigma_2^2$.

iii. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\mid\delta)>\alpha_0$，当 $\sigma_1^2>\sigma_2^2$.

iv. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\mid\delta)\to 0$，当 $\sigma_1^2/\sigma_2^2\to 0$.

v. $\pi(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\mid\delta)\to 1$，当 $\sigma_1^2/\sigma_2^2\to\infty$.

当 $\sigma_1^2>\sigma_2^2$ 时，用类似于式(8.5.11)的讨论可以说明 $\Pr(V\geqslant c)>\alpha_0$，而与 μ_1 和 μ_2 的值无关. 该讨论也证明了，当 $\sigma_1^2/\sigma_2^2\to 0$ 时，$\Pr(V\geqslant c)\to 0$；当 $\sigma_1^2/\sigma_2^2\to\infty$ 时，$\Pr(V\geqslant c)\to 1$.

大多数统计包可以算出 F 分布的密度函数和或分位数. 密度函数可以用来计算我们刚才导出的检验的功效函数. 当 $V\geqslant c$ 时，拒绝 H_0: $\sigma_1^2\leqslant\sigma_2^2$ 的检验 δ 的功效函数是

$$\pi(\sigma_1^2,\sigma_2^2,\mid\delta)=\Pr(V\geqslant c\mid\sigma_1^2,\sigma_2^2)=\Pr\left(V^*\geqslant c\frac{\sigma_2^2}{\sigma_1^2}\mid\sigma_1^2,\sigma_2^2\right)$$

$$=1-F_{m-1,n-1}\left(c\frac{\sigma_2^2}{\sigma_1^2}\right),$$

此处 $F_{m-1,n-1}$ 是自由度为 $m-1$，$n-1$ 的 F 分布的密度函数，例如在例 8.5.2 中，当 $\sigma_1^2=3\sigma_2^2$ 时，若想计算水平为 0.05 时的功效，使用计算机程序算出

$$1-F_{5,20}\left(2.71\times\frac{1}{3}\right)=0.498.$$

即使当 σ_1^2 是 σ_2^2 的 3 倍时，在 0.05 水平下，拒绝 H_0 的概率大约只有 50%. [510]

8.5.6 双边备择

假定我们要检验这样的假设：

$$H_0:\sigma_1^2=\sigma_2^2,\qquad\qquad (8.5.12)$$
$$H_1:\sigma_1^2\neq\sigma_2^2.$$

如果 $V\leqslant c_1$ 或 $V\geqslant c_2$，那么拒绝 H_0，应该是合理的. V 在例 8.5.9 中已经定义了，c_1 和 c_2 是常数且满足当 $\sigma_1^2=\sigma_2^2$ 时，$\Pr(V\leqslant c_1)+\Pr(V\geqslant c_2)=\alpha_0$. c_1 和 c_2 最方便的选择是满足 $\Pr(V\leqslant c_1)=\Pr(V\geqslant c_2)=\alpha_0/2$，即选择 c_1 和 c_2 分别为适当的 F 分布的 $\alpha_0/2$ 分位数和 $1-\alpha_0/2$ 分位数.

例 8.5.3 来自人工降雨云中的雨 在 8.4 节的习题 1，我们比较了人工云层和非人工云层的降雨量的平均值，假设的前提是两个方差相等. 现在检验两个方差相等的零假设，备择假设是两方差不相等，显著性水平 $\alpha_0=0.05$. 利用在该习题中所给的统计数值，因为 $m=n$，V 的值是 63.96/67.39 = 0.9491. 需要把该值与自由度为 (25,25) 的 F 分布的分位数 0.025 和 0.975 做比较. 因为 F 分布的分位数表格没有对应自由度为 25 的行或列，可以做 20 ~ 30 之间的自由度内插或使用电脑程序来计算这些分位数. 这两个分位数的值是 0.4484 和 2.2303. 因为 V 在这两个数字之间，在 $\alpha_0=0.05$ 的水平下要接受原假设. ◀

当 $m\neq n$，就不可能借助某种统计量 T 把上述的双边 F 检验写成“如果 $T\geqslant c$ 就

拒绝原假设”的形式. 由于该原因，我们将不对双边检验计算 p-值(见习题15).

8.5.7 扩展研究：F 检验作为似然比检验

在本部分将说明假设(8.5.8)的 F 检验是似然比检验. 在观察到两个样本的值 $x_1,\cdots,x_m$ 及 $y_1,\cdots,y_n$ 之后，似然函数 $g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$ 就是

$$g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)=f_m(\boldsymbol{x}|\mu_1,\sigma_1^2)f_n(\boldsymbol{y}|\mu_2,\sigma_2^2).$$

这里 $f_m(\boldsymbol{x}|\mu_1,\sigma_1^2)$ 和 $f_n(\boldsymbol{y}|\mu_2,\sigma_2^2)$ 具有在式(8.3.10)中所给的一般形式. 对(8.5.8)中的假设，Ω_0 包含所有满足 $\sigma_1^2\leqslant\sigma_2^2$ 的参数 $\theta=(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$，其中，$\Omega_1$ 包含所有满足 $\sigma_1^2>\sigma_2^2$ 的 θ，似然比是

$$\Lambda(\boldsymbol{x},\boldsymbol{y})=\frac{\sup_{\{(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)\ |\ \sigma_1^2>\sigma_2^2\}}g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}{\sup_{\{(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)\ |\ \sigma_1^2\leqslant\sigma_2^2\}}g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}.\tag{8.5.13}$$

然而该似然比检验指明：如果 $\Lambda(\boldsymbol{x},\boldsymbol{y})\geqslant k$，那么 H_0 被拒绝，通常由检验所希望的水平 α_0 来选择 k.

511

为了便于(8.5.13)中的最大值的计算，令

$$s_x^2=\sum_{i=1}^m(x_i-\bar{x}_m)^2 \text{ 和 } s_y^2=\sum_{i=1}^n(y_i-\bar{y}_n)^2.$$

然后记

$$g(\boldsymbol{x},\boldsymbol{y}|\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$$
$$=\frac{1}{(2\pi)^{(m+n)/2}\sigma_1^m\sigma_2^n}\exp\left(-\frac{1}{2\sigma_1^2}[n(\bar{x}_m-\mu_1)^2+s_x^2]-\frac{1}{2\sigma_2^2}[n(\bar{y}_n-\mu_2)^2+s_y^2]\right).$$

对于(8.5.13)中的分子和分母，需要 $\mu_1=\bar{x}_m$ 和 $\mu_2=\bar{y}_n$ 来最大化这个似然值. 如果 $s_x^2/m>s_y^2/n$，那么分子在条件 $\sigma_1^2=s_x^2/m$ 和 $\sigma_2^2=s_y^2/n$ 时最大化了. 如果 $s_x^2/m\leqslant s_y^2/n$，那么这些值也能使分母最大化. 对于另外两种情况(分子在 $s_x^2/m\leqslant s_y^2/n$ 时和分母在 $s_x^2/m>s_y^2/n$ 时)，可以直接证明：为了实现最大化要求 $\sigma_1^2=\sigma_2^2$. 在这些情况下，最大化发生在

$$\sigma_1^2=\sigma_2^2=\frac{s_x^2+s_y^2}{m+n}.$$

把所有这些值代入(8.5.13)得

$$\Lambda(\boldsymbol{x},\boldsymbol{y})=\begin{cases}dw^{m/2}(1-w)^{n/2} & \text{若 } s_x^2/m\leqslant s_y^2/n,\\ (dw^{m/2}(1-w)^{n/2})^{-1} & \text{若 } s_x^2/m>s_y^2/n,\end{cases}$$

其中

$$w=\frac{s_x^2}{s_x^2+s_y^2} \text{ 和 } d=\frac{(m+n)^{(m+n)/2}}{m^{m/2}n^{n/2}}.$$

注意，当且仅当 $w\leqslant m/(m+n)$ 时有 $s_x^2/m\leqslant s_y^2/n$，其次利用以下事实：当 $0<w<m/(m+n)$ 时函数 $h(w)=w^{m/2}(1-w)^{n/2}$ 是递增的，当 $m/(m+n)<w<1$ 时它是递减的.

最后，注意 $h(m/[m+n])=1/d$. 由此，$\Lambda(\boldsymbol{x},\boldsymbol{y})\geqslant k$ 成立当且仅当 $w\geqslant k'$ 对于另外某个常数 k' 成立，进而等价于 $s_x^2/s_y^2\geqslant k''$. 因为 s_x^2/s_y^2 是观察值 V 乘以一个正的常数，所以似然比检验当 V 很大时拒绝 H_0. 这与 F 检验一样.

在假设中的不等式取相反方向的情况下，大家可以容易地修改以上的讨论. 当假设是(8.5.12)，检验是双边的. 可以证明(见习题16)，显著性水平 α_0 的似然比检验在 $V\leqslant c_1$ 或 $V\geqslant c_2$ 时拒绝 H_0. 不幸的是，通常计算出必需的 c_1 和 c_2 很不令人满意的. 由于该原因，在这里人们经常放弃严格的似然比标准而是简单地令 c_1、c_2 分别为适当的 F 分布的 $\alpha_0/2$ 和 $1-\alpha_0/2$ 分位数. 512

8.5.8 小结

如果 Y 和 W 是互相独立的，并且 Y 服从自由度为 m 的 χ^2 分布，W 服从自由度为 n 的 χ^2 分布，则 $(Y/m)/(W/n)$ 服从自由度为 (m,n) 的 F 分布，假设我们观察来自于两个可能有不同方差的正态分布的两个独立随机样本，当两个方差相等时，两个方差的无偏估计的比率 V 服从 F 分布. 通过把 V 和 F 分布的各种分位数相比较，可以构造出关于两个方差的假设检验.

习题

1. 继续考虑8.4节第11题所描述的情况. 原假设为看过物体图片的实验对象所需合成时间的方差不小于未看过图片的实验对象相应的方差. 而备择假设则是前一方差小于后一方差. 用0.05的显著性水平.
2. 假设一个随机变量 X 服从自由度为3和8的 F 分布，确定 c 的值，使其满足 $\Pr(X>c)=0.975$.
3. 假设随机变量 X 服从自由度为1和8的 F 分布，用 t 分布表来确定 c 的值使 $\Pr(X>c)=0.3$ 成立.
4. 假设随机变量 X 服从自由度为 m，n 的 F 分布 $(n>2)$，证明：$E(X)=n/(n-2)$. 提示：若 Z 服从自由度为 n 的 χ^2 分布，求出 $E(1/Z)$ 的值.
5. 当 $m=n$ 时，一个自由度为 m，n 的 F 分布的中位数的值是多少?
6. 假设随机变量 X 服从自由度为 m，n 的 F 分布，证明：随机变量 $mX/(mX+n)$ 服从参数为 $\alpha=m/2$ 和 $\beta=n/2$ 的 β 分布.
7. 考虑两个不同的正态分布，它们的均值 μ_1 和 μ_2，以及方差 σ_1^2 和 σ_2^2 都是未知的. 如果要检验以下假设:

$$H_0: \sigma_1^2\leqslant\sigma_2^2,$$
$$H_1: \sigma_1^2>\sigma_2^2.$$

进一步假定有一个随机样本包含来自于第一个正态分布的16个观察值，已知 $\sum_{i=1}^{16}X_i=84$ 和 $\sum_{i=1}^{16}X_i^2=563$，而另一个独立的随机样本包含来自于第二个正态分布的10个观察值. 已知 $\sum_{i=1}^{10}Y_i=18$ 和 $\sum_{i=1}^{10}Y_i^2=72$.

a. 求 σ_1^2 和 σ_2^2 的 M. L. E. .

b. 在 0.05 的显著水平下做 F 检验. 问假设 H_0 是否被拒绝?

8. 继续考虑习题 7，但是改用下列假设

$$H_0: \sigma_1^2 \leqslant 3\sigma_2^2,$$
$$H_1: \sigma_1^2 > 3\sigma_2^2.$$

描述怎样用 F 检验来检验这些假设?

9. 继续考虑习题 7 的情况，但是现在考虑下列假设

$$H_0: \sigma_1^2 = \sigma_2^2,$$
$$H_1: \sigma_1^2 \neq \sigma_2^2.$$

设统计量 V 由(8.5.9)式定义. 我们希望当 $V \leqslant c_1$ 或者 $V \geqslant c_2$ 拒绝 H_0，其中常数 c_1 和 c_2 的选择要使当 H_0 为真时，$\Pr(V \leqslant c_1) = \Pr(V \geqslant c_2) = 0.025$. 当 $m = 16$ 和 $n = 10$ 时，像习题 7 一样，确定 c_1 和 c_2 的值.

10. 假定一个随机样本包含 16 个观察值，取自一个均值 μ_1 和方差 σ_1^2 均未知的正态分布. 另一个独立的随机样本，包含 10 个观察值，取自均值 μ_2 和方差 σ_2^2 均未知的另一个正态分布. 对于每个常数 $r > 0$，在显著性水平为 0.05 时，使用下列假设

513

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = r,$$

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq r.$$

构造一个检验.

11. 再考虑习题 10 的情况，用该习题的结果来计算 σ_1^2/σ_2^2 的置信区间（置信系数为 0.95).

12. 假设一个随机变量 Y 服从自由度为 m_0 的 χ^2 分布，取常数 c 使 $\Pr(Y > c) = 0.05$，解释为什么在 F 分布的 0.95 的分位数的表中，当 $m = m_0$ 和 $n = \infty$ 时，表值等于 c/m_0.

13. 在 F 分布 95% 的分位数的表格中的最后一列包含当 $m = \infty$ 时的值，解释怎样从 χ^2 分布表中得到这一栏的表值.

14. 继续考虑习题 7 的情况，当 $\sigma_1^2 = 2\sigma_2^2$ 时，找出 F 检验的功效函数.

15. 双边 F 检验的 p-值定义有如下另外一种方法：如果 $V = v$ 是观察值，这里 V 由等式(8.5.9)定义，则 p-值为

$$\min\{2[1 - F_{m-1,n-1}(v)], 2F_{m-1,n-1}(v)\}. \tag{8.5.14}$$

a. 证明：在(8.5.14)中的两个数字一个至少为 1，另一个至多为 1;

b. 用这个定义，计算例 8.5.3 中的 p-值.

16. 设 V 由(8.5.9)式定义，要确定(8.5.12)假设的 α_0 水平似然比检验，证明：当 $V \leqslant c_1$ 或 $V \geqslant c_2$ 时，似然比检验将拒绝 H_0，其中当 $\sigma_1^2 = \sigma_2^2$ 时，$\Pr(V \leqslant c_1) + \Pr(V \geqslant c_2) = \alpha_0$.

514

17. 证明在习题 9 中找到的检验不是似然比检验.

第 9 章　分类数据与非参数方法

9.1　拟合优度检验

在一些问题中，对将要观察的数据我们在心目中已有了一个明确的分布. 如果该分布不是很合适，那么不需要知道可供选择的分布的参数族. 在这种情况下或其他情况下，我们仍能检验来自于某一特定分布的数据的原假设及其备择假设，即数据不是来自这一分布.

9.1.1　非参数问题的描述

在第 6、第 7 和第 8 章中我们所提到的有关估计和假设检验的每个问题中，假定统计员所得到的观测值来自于形式已知的分布，尽管一些参数的值未知. 例如，可能假定组成随机样本的观察值来自于一个均值未知的泊松分布，或可能假定观察值来自于均值和方差都未知的两参数正态分布. 换言之，我们已经假定观察值来自于某个特定分布的参数族，从而得到参数值的统计推断来确定该分布族.

在本章要讨论的许多问题中，不需要假定得到的观测值来自于某个特定分布的参数族. 而是要在不做关于分布形式的任何假设的情况下，研究得到观察值所来自的某分布的推断. 举一个例子，我们可以简单地假设组成随机样本的观察值来自于一个连续型分布的随机样本，但不再进一步作这种分布形式的具体假定，接下来可以研究该种分布是正态分布的可能性. 第二个例子中，我们也许有意对于抽取样本的某个分布的中位数的值作推断，而这里仅仅假设这种分布是连续的. 第三个例子中，我们也许有兴趣研究两个独立的随机样本来自于同一种分布的概率，而仅仅假设从中抽取样本的两个分布都是连续的.

观察值的可能分布不限制在某个指定参数族的一类问题称为*非参数问题*，而用于处理该类问题的统计方法称为*非参数统计方法*.

9.1.2　分类数据

在本节和接下来的四节中我们将考虑基于以下的数据的统计问题，可以将数据中每个观察值分类属于有限多个可能类型中的一个. 称此类型的观察值为*分类数据*. 因为在这些问题中只包含了有限多个可能的类型，而且因为对于该种类型的概率作推断感兴趣，那么这些问题实际上只涉及到有限多个参数. 然而，正如我们将要看到的，基于分类数据的统计方法可以很有用地运用到参数和非参数的问题中.

9.1.3 χ^2 检验

假设总体由 k 个不同类型的个体组成，令 p_i 表示属于类型 $i(i=1,\cdots,k)$ 的一个个体被随机选中的概率．当然，$p_i>0, i=1,\cdots,k$ 且 $\sum_{i=1}^{k}p_i=1$. 设 $p_1^0,\cdots,p_k^0$,是满足 $p_i^0>0$的一组特定的数．这里 $i=1,\cdots,k$ 且 $\sum_{i=1}^{k}p_i^0=1$．假定要检验下面的假设：

$$
\begin{aligned}
&H_0: p_i=p_i^0, \quad \text{对于 } i=1,\cdots,k,\\
&H_1: p_i\neq p_i^0, \quad \text{至少有一个 } i \text{ 满足}.
\end{aligned}
\tag{9.1.1}
$$

我们假定从给定的总体中取出容量为 n 的随机样本．即，取出 n 个独立的观察值，而且每个观察值属于类型 $i(i=1,\cdots,k)$的概率为 p_i．基于这 n 个观测值，要检验假设(9.1.1).

对于$i=1,\cdots,k$,设 N_i 表示随机样本中属于类型i的观察值的个数．因此，$N_1,\cdots,N_k$ 是满足 $\sum_{i=1}^{k}N_i=n$ 的非负整数．事实上，$(N_1,\cdots,N_n)$是服从参数为 n 和 $\boldsymbol{p}=(p_1,\cdots,p_k)$ 的多项分布．当原假设 H_0 为真时，属于第i类观察值的期望值为 $np_i(i=1,\cdots,k)$. 观察值的实际值 N_i 和期望值 np_i 的差距在 H_0 为真时要比当 H_0 不为真时要小．这看上去很合理，因此，要对这个误差 $N_i-np_i^0(i=1,\cdots,k)$的值建立一个假设检验，当这些误差的量相当大时拒绝 H_0.

536

在 1900 年，Karl Pearson 提出使用下面的统计量：

$$
Q=\sum_{i=1}^{k}\frac{(N_i-np_i^0)^2}{np_i^0}. \tag{9.1.2}
$$

此外，Pearson 还证明了如果假设 H_0 为真，那么当样本容量 $n\to\infty$ 时，Q 的密度函数趋近于自由度为 $k-1$ 的χ^2 分布的密度函数．因此，如果 H_0 为真，且样本容量 n 足够大，Q 的分布将近似于自由度为 $k-1$ 的χ^2 分布．这里做的讨论表示当 $Q\geqslant c$ 时应该拒绝 H_0，其中 c 取适当的常数．如果要在显著性水平为 α_0 下进行检验，那么 c 应该取自由度为 $k-1$ 的χ^2 分布的 $1-\alpha_0$ 分位数．该种检验被称作为χ^2 拟合优度检验.

注：χ^2 检验统计量的一般形式．(9.1.2)中统计量 Q 的形式对于所有χ^2 检验包括那些在本章后要介绍的检验都适用．这种形式是一些项的和，其中每一项都是观察数和期望数的差值的平方除以期望数：$\sum$(观察数－期望数$)^2$/期望数．期望数都是假定原假设为真时计算得出的.

只要每个期望值 np_i^0 的值不太小，χ^2 分布将会是 Q 真实分布的一个很好的近似．特别地，如果 $np_i^0\geqslant 5$,其中 $i=1,\cdots,k$，该近似程度将会很好，而且如果 $np_i^0\geqslant 1.5$，当 $i=1,\cdots,k$ 时，该近似程度仍然会很令人满意.

现在通过考虑两个例子来说明χ^2 拟合优度检验的用法.

9.1.4 检验关于比例的假设

假定在一个大样本里次品的比例 p 是未知的，检验下面的假设：

$$
\begin{aligned}
&H_0: p = 0.1, \\
&H_1: p \neq 0.1.
\end{aligned} \tag{9.1.3}
$$

又设在一个100个产品的随机样本中，发现16个是有缺陷的. 要通过χ^2拟合优度检验来检验假设(9.1.3).

由于在本例中只有两种类型的产品，即次品和正品，且我们已知$k=2$. 此外，如果我们令p_1表示次品的未知比例，p_2表示正品的未知比例，那么可以将假设(9.1.3)表示成为下面的形式：

$$
\begin{aligned}
&H_0: p_1 = 0.1 \text{ 且 } p_2 = 0.9, \\
&H_1: \text{假设 } H_0 \text{ 是错的.}
\end{aligned} \tag{9.1.4}
$$

当样本容量$n=100$，若H_0是真时，次品数的期望值是$np_1^0=10$，正品的期望值为$np_2^0=90$. 令N_1表示样本中次品的个数，N_2表示样本中正品的个数. 那么，当H_0为真时，由等式(9.1.2)所定义的统计量Q的分布近似于自由度为1的χ^2分布. [537]

在本例中，N_1等于16且N_2等于84，且可以发现Q的值为4. 这可以由自由度为1的χ^2分布表来决定，即与$Q=4$相对应的落在0.025与0.05之间的尾部面积相一致. 因此，在置信度为$\alpha_0=0.05$的显著性水平下，原假设H_0被拒绝，但是在水平$\alpha_0=0.025$时，H_0会被接受. 关于单个比例的假设，我们已在第8章中给出了检验(例如，参见8.1节习题11).

例9.1.1 蒙大拿州的民意调查 蒙大拿大学的商业和经济研究处在1992年5月对蒙大拿州的居民进行了民意调查评估. 应答者被问及他们的个人经济地位是更差了、相同，还是比上年更好. 结果见表9-1.

表9-1 美国蒙大拿州关于个人经济地位问题的民意调查

更差	相同	更好	总和
58	64	67	189

我们可能对应答者对3种可能答案的回答是否服从均匀分布有兴趣. 也就是说，可以检验：3种答案的可能性都等于1/3为原假设. 计算

$$
Q = \frac{(58-189/3)^2}{189/3} + \frac{(64-189/3)^2}{189/3} + \frac{(67-189/3)^2}{189/3} = 0.6667.
$$

由于0.6667是自由度为2的χ^2分布的0.283分位点. 我们只能在比$1-0.283=0.717$的水平更高时才拒绝原假设. ◀

9.1.5 连续分布的假设检验

考虑一个随机变量X在区间$0<X<1$上取值，但是在此区间上有一个未知的概率密度函数. 假设从该未知分布中抽取一个有100个观察值的随机样本. 希望检验原假设：未知分布是区间[0,1]上的均匀分布，而其相应的备择假设是分布是非均匀的. 该问题是一个非参数问题，因为X的分布可能是区间[0,1]上任意一个连续分

布. 然而，我们现在要证明的是χ^2拟合优度检验可以被应用在该问题上.

假如把区间[0，1]分成20个长度相等的子区间，即[0,0.05),[0.05,0.1),…. 如果实际的分布是一个均匀分布，那么每个观察值落在第i个子区间内的概率为1/20，其中$i=1,\cdots,20$. 既然在本例中的样本容量n为100，则观察值落入每个子区间内的期望值为5. 若N_i表示为在样本中属于第i个子区间的观察值的个数，那么由等式(9.1.2)所定义的统计量Q可以重新简化为如下形式：

538

$$Q=\frac{1}{5}\sum_{i=1}^{20}(N_i-5)^2. \tag{9.1.5}$$

若原假设是真的且所取观察值的分布确实是一个均匀分布. 那么Q将近似于一个自由度为19的χ^2分布.

在本例中所用的方法明显可以被用到每一个连续分布，为了检验观察值的一个随机样本是否来自于一个特殊分布，可以运用以下的步骤：

i. 把整条实数轴或任意概率为1的特定区间分割成有限的k个不相交的子区间. 通常所选的k要满足当H_0为真时，落在每个子区间的观察值的期望值至少为5.

ii. 确定特定的假设分布分配到第i个子区间的概率p_i^0，并计算落在第i个子区间内观察值的期望数np_i^0.

iii. 计算样本中落入第i个子区间内的样本观察值的个数$N_i(i=1,\cdots,k)$.

iv. 计算由等式(9.1.2)所定义的Q的值. 若假设的分布是正确的，则Q将近似服从于一个自由度为$k-1$的χ^2分布.

例9.1.2　滚珠轴承的失效次数　在例5.6.7中我们观察到23个滚珠轴承的失效次数，把这些失效时间的对数看成是正态随机变量. 假定希望用χ^2检验来检验原假设，即其寿命时间的对数值是来自于均值为$\ln(50)=3.912$，方差为0.25的正态分布的独立随机样本. 为了在每个区间中的期望数都至少为5，我们至多可用$k=4$个区间. 在原假设条件下，使每个区间的概率为0.25. 也就是说对所假设的正态分布在0.25，0.5和0.75分位数时分成四个区间，这些分位数的值为：

$$\begin{aligned}
3.912+0.5\Phi^{-1}(0.25) &= 3.192+0.5\times(-0.674) &&= 3.575,\\
3.912+0.5\Phi^{-1}(0.5) &= 3.192+0.5\times 0 &&= 3.912,\\
3.912+0.5\Phi^{-1}(0.75) &= 3.192+0.5\times 0.674 &&= 4.249,
\end{aligned}$$

它在标准正态分布下0.25，0.75的分位数分别为±0.674. 观察到的对数值分别为：

2.88	3.36	3.50	3.73	3.74	3.82	3.88	3.95
3.95	3.99	4.02	4.22	4.23	4.23	4.23	4.43
4.53	4.59	4.66	4.66	4.85	4.85	5.16	

539

在这四个区间中每个区间内的观测值的数目为3，4，8和8. 然后计算

$$Q=\frac{(3-23\times0.25)^2}{23\times0.25}+\frac{(4-23\times0.25)^2}{23\times0.25}+\frac{(8-23\times0.25)^2}{23\times0.25}+\frac{(8-23\times0.25)^2}{23\times0.25}=3.609.$$

在χ^2分布表中当自由度为3时指出3.609是在0.6与0.7分位数之间，因此在小于0.3的水平下将接受原假设，在大于0.4水平下拒绝原假设（实际上，p-值为0.307）. ◀

在刚刚描述的过程有一个带有主观性特征的是选择子区间的方法. 对于同一个问题，两个数据分析员可能用不同的方法来很好地选择子区间. 一般来说，选取子区间使得落在每个子区间的观察值的期望数近似相等，并且，在保证每个区间内的期望值不变小的前提下，选择尽可能多的区间个数，这是一个很好的准则. 正如我们在例9.1.2中已做的那样.

9.1.6 扩展研究：检验过程的讨论

拟合优度的χ^2检验法受到一些批判. 特别是，在χ^2检验中的原假设H_0仅仅指定了观察值的分布，但有可能观察值的实际分布与来自于该指定分布的一个随机样本的分布恰好相同. 因此，如果χ^2检验是基于大量的观察值，我们几乎可以肯定与Q的观察值相对应的尾部面积会很小. 由于这个原因，如果没有进一步分析，一个很小的尾部面积不应该被视为强有力的理由而拒绝假设H_0. 在一个数据分析员推断假设H_0是不合理之前，他应该肯定存在合理的备择分布使得观察值拟合程度更好. 例如，一个数据分析员为了肯定至少存在一个备择分布，其尾部面积相对应的Q的计算值比H_0所指定的分布要大得多，而有可能要对一些合理的备择分布来计算统计量Q的值.

χ^2拟合优度检验的一个特别之处是其过程被设计成检验原假设H_0: $p_i = p_i^0$，当$i=1,\cdots,k$，对应的备择假设是H_0不真. 如果希望用一个特别有效的检验步骤来检测出假定值$p_1^0,\cdots,p_k^0$和真实值$p_1,\cdots,p_k$的方差的某种类型，那么数据分析员应该设计特殊的检验使得对于功效高的备择假设比功效低的备择假设兴趣少些. 这个话题在此书中不作讨论.

540

因为在方程式(9.1.2)中的随机变量$N_1,\cdots,N_k$是离散的，χ^2与Q分布的近似程度，有时候可以通过引进一个在5.8节中描述过的连续性修正而得到改善，然而，此书中我们不用这种修正.

9.1.7 小结

已经介绍了χ^2拟合优度检验是用来检验数据来自于某一确定分布，形成一个独立同分布的样本这一原假设与其备择假设，即数据来自于其他分布的一种检验方法. 当指定分布是离散时该检验是很正常的. 假设对于每次观察都有k个可能的值，且观察到有N_i个取值为i, $i=1,\cdots,k$. 假定原假设中认为第i个可能取值的概率为p_i^0，其中$i=1,\cdots,k$. 那么计算得出

$$Q=\sum_{i=1}^{k}\frac{(N_i-np_i^0)^2}{np_i^0},$$

此处 $n=\sum_{i=1}^{k}N_i$ 是样本容量. 当原假设认为数据服从连续分布时，那么首先必须建立一个与其相对应的离散分布. 通过将实数轴划分成有限多个区间(记为 k 个)，并计算每个区间的概率 $p_1^0,\cdots,p_k^0$，然后假定我们从数据中获知每一个观察值落入了哪一个区间. 这就把原始数据转变成为 k 个可能的离散数据. 例如，在对 Q 的公式中使用的 N_i 的值是落入第 i 个区间的观察值的个数. 在本文中所有 χ^2 检验统计量都有 Σ(观察值 - 期望值)^2/期望值的形式，此处的"观察值"代表被观察到的个数，"期望值"代表在原假设是真的前提下被观察到的个数的期望值.

习题

1. 考虑检验假设(9.1.4)的一个数值例子. 使用8.1节中习题11中指出的检验步骤，并把结果与本节中所获得的数值结果进行比较.
2. 证明：如果 $p_i^0=1/k$,其中 $i=1,\cdots,k$,那么由(9.1.2)式所定义的统计量 Q 可以写成如下形式
$$Q=\left(\frac{k}{n}\sum_{i=1}^{k}N_i^2\right)-n.$$
3. 调查你最喜欢的伪随机数产生器的随机性. 模拟200个从0到1之间的伪随机数，并把单位区间分割成每个长度为0.1的 $k=10$ 个区间. 对于以下假设使用 χ^2 检验：假设10个区间中的每个区间都有相同的包含一个伪随机数据的概率.
4. 根据简单基因原则，如果一个孩子的母亲和父亲的基因类型都是Aa，那么这个孩子将会有1/4的可能性是基因类型为AA，有1/2的可能性是基因类型为Aa，1/4的可能性是基因类型为aa. 对于父母基因类型都是Aa的24个孩子的一个随机样本中，发现其中10个孩子的基因类型是AA，10个孩子的基因类型是Aa，4个孩子的基因类型是aa. 试运用 χ^2 拟合优度检验来研究简单基因原则是否正确.

541

5. 设在一个 n 重伯努利试验中，每次试验成功的概率 p 是未知的. 又设 p_0 是在区间(0,1)中一个给定的数，并且希望检验如下的假设：
$$H_0: p=p_0,$$
$$H_1: p\neq p_0.$$
令 $\overline{X}_n$ 表示在 n 次试验中成功的比例，并假定使用 χ^2 拟合优度检验来检验上述所给的假设.
   a. 证明：由式(9.1.2)中所定义的统计量 Q 能写成如下形式
$$Q=\frac{n(\overline{X}_n-p_0)^2}{p_0(1-p_0)}.$$
   b. 假定 H_0 为真，证明：当 $n\to\infty$ 时，Q 的密度函数收敛于自由度为1的 χ^2 分布的密度函数. 提示：证明 $Q=Z^2$，由中心极限定理可知，Z 是一个随机变量，它的密度函数收敛于标准正态分布的密度函数.
6. 由一个标准工序生产的30%小钢棒当遭受一个3000磅的负荷物时会损坏. 在一个由一种新工序生产的50根相似钢棒所组成的样本中，发现当遭受3000磅的负荷物时，有21个损坏了. 运用 χ^2 拟合优度检验来检验新工序的损坏率与旧工序的损坏率是一样的这一

假设.

7. 随机地从区间(0,1)中抽取1800个观察值的作为样本，发现有391个值是在0到0.2之间，490个值是在0.2到0.5之间，580个值是在0.5到0.8之间，以及339个值是在0.8到1之间. 在0.01的显著性水平下运用χ^2拟合优度检验来检验假设：随机样本是来自于[0,1]区间上的均匀分布.

8. 设居住在同一大城市中男人身高的分布是一个正态分布，它的均值为68英寸，标准差为1英寸. 假如测量了居住在与该城市相近的地区里500个男人的身高，得到了表9-2的分布. 检验以下假设：就身高而言，这500个男人所组成的这一随机样本是来自于居住在该城市的男人.

表9-2 习题8的数据

身高	小于等于66	在66与67.5之间	在67.5与68.5之间	在68.5与70之间	大于等于70
男人的数量	18	177	198	102	5

9. 在表9-3中有50个数据是来自于某一标准正态分布的随机样本.

表9-3 习题9的数据

-1.28	-0.32	-1.38	2.33	-1.14	0.41	-0.49	1.05	0.35	0.64
-1.22	-0.80	-1.26	-0.34	0.64	-0.01	0.36	0.04	2.82	0.56
-0.45	-1.66	0.49	-1.96	3.44	0.67	-1.24	0.76	-0.46	-0.11
-0.35	1.39	-0.14	-0.64	-1.67	-1.13	-0.04	0.61	-0.63	0.13
0.72	0.38	-0.85	-1.32	0.85	-0.41	-0.11	-2.04	-1.61	-1.81

a. 把实数轴分为5个区间，使得在标准正态分布下每个区间有0.2的概率，作χ^2拟合优度检验.

b. 把实数轴分为10个区间，使得在标准正态分布下每个区间有0.1的概率，作χ^2拟合优度检验.

9.2 复合假设的拟合优度

可以推广拟合优度检验来处理一些数据分布属于一个特殊参数族的原假设. 备择假设为数据服从的分布不属于该参数族. 简单原假设与复合原假设检验过程有两个不同. 第一，在检验中统计量Q，概率p_i^0被基于参数族估计所得的概率代替. 第二，自由度因参数个数而减少.

542

9.2.1 复合原假设

可以再考虑包含k个不同种类项的一个总体，并用p_i来表示随机抽取的每一个项类型为$i(i=1,\cdots,k)$的概率. 现在假定要检验的不是参数$p_1,\cdots,p_k$有指定取值的原假设，而是对检验$p_1,\cdots,p_k$的值属于可能取值的某个指定子集的复合原假设感兴趣. 特别地，要考虑这样的问题：原假设中指定的参数$p_1,\cdots,p_k$实际上可以表示成

少量参数的函数形式.

比如，考虑一个基因问题，该问题中一个基因有2个不同的染色体(参见例1.6.3)，一个给定的总体中的每一个个体都必须有三种可能基因类型中的一种. 如果从父母那里继承染色体是独立的，且每对父母将第一染色体传给子女的概率是相同的，那么三种不同基因类型的概率 p_1，p_2 和 p_3 可以用以下形式表示：

$$p_1=\theta^2, p_2=2\theta(1-\theta), p_3=(1-\theta)^2. \tag{9.2.1}$$

在此，参数 θ 值未知且可取区间 $0<\theta<1$ 中的任意值. 对这区间上的每个 θ 值，可以看到当 $i=1,2$ 或3时，$p_i>0$，且 $p_1+p_2+p_3=1$. 在这个问题中，从总体中随机抽出一个随机样本，统计者必须用每种基因拥有者的观察数，来决定是否有理由相信对于区间 $0<\theta<1$ 中某个 θ 值，p_1，p_2 和 p_3 具有(9.2.1)式所假设的形式.

如果一个基因有三个不同的染色体，总体中的每个个体都必有6种可能的基因类型中的一种. 同样地，若从父母那儿继承染色体是独立的，且假定每对父母将第一及第二染色体传给子女的概率各为 θ_1 和 θ_2，则不同类型基因的概率 $p_1,\cdots,p_6$ 可以用下面的形式来表示，当 θ_1 和 θ_2 的取值满足 $\theta_1>0,\theta_2>0$，且 $\theta_1+\theta_2<1$ 时：

$$\begin{aligned}&p_1=\theta_1^2, p_2=\theta_2^2, p_3=(1-\theta_1-\theta_2)^2, p_4=2\theta_1\theta_2,\\&p_5=2\theta_1(1-\theta_1-\theta_2), p_6=2\theta_2(1-\theta_1-\theta_2).\end{aligned} \tag{9.2.2}$$

同样地，对于所有符合上述条件的 θ_1 和 θ_2 的取值，可以证明对于 $i=1,\cdots,6$，有 $p_i>0$ 且 $\sum_{i=1}^{6}p_i=1$. 基于一个随机样本中拥有每种基因个体的观察数值 $N_1,\cdots,N_6$，统计者必须决定到底是接受还是拒绝以下原假设：对 θ_1 和 θ_2 的某些值，概率 $p_1,\cdots,p_6$ 可以表示成(9.2.2)式的形式.

543

从形式上看，对于刚刚所考虑的该类问题，我们感兴趣的是检验假设：对于 $i=1,\cdots,k$，每个概率 p_i 可以被表示成一参数矢量 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_s)$ 所构成的特殊函数 $\pi_i(\boldsymbol{\theta})$. 假设 $s<k-1$ 且矢量 $\boldsymbol{\theta}$ 没有分量可以被表示成其他 $s-1$ 个分量的函数形式. 可以用 Ω 来表示所有 $\boldsymbol{\theta}$ 的可能值所构成 的 s 维参数空间. 此外还假设在以下意义下函数式 $\pi_1(\boldsymbol{\theta}),\cdots,\pi_k(\boldsymbol{\theta})$ 总可以形成 $p_1,\cdots,p_k$ 值的一个可行的集合：对于每个 $\boldsymbol{\theta}\in\Omega$ 的值，和每个 $i=1,\cdots,k$ 都有 $\pi_i(\boldsymbol{\theta})>0$，且 $\sum_{i=1}^{k}\pi_i(\boldsymbol{\theta})=1$.

被检验的假设可以表示成以下的形式：

$$\begin{aligned}&H_0:\text{对于 } i=1,\cdots,k,\text{存在一个值 } \boldsymbol{\theta}\in\Omega,\text{使得 } p_i=\pi_i(\boldsymbol{\theta}),\\&H_1:H_0\text{ 假设不正确.}\end{aligned} \tag{9.2.3}$$

$s<k-1$ 的假定保证了假设 H_0 事实上把 $p_1,\cdots,p_k$ 的取值限制在这些概率的所有可能取值集合的一个真子集上. 即，当向量 $\boldsymbol{\theta}$ 取遍集合 Ω 中所有值时，向量 $[\pi_1(\boldsymbol{\theta}),\cdots,\pi_k(\boldsymbol{\theta})]$ 仅取遍 $(p_1,\cdots,p_k)$ 可能取值的一个真子集.

9.2.2 复合原假设的 χ^2 检验

为了进行假设(9.2.3)拟合优度的 χ^2 检验，必须修改式(9.1.2)中定义的统计量

Q，因为在一个有 n 个观测值的随机样本中，类型 i 的观察值的期望 np_i^0 不再完全由原假设 H_0 所确定. 在 H_0 为真的假设下，要做的修正只须简单地用这个期望值的极大似然估计来取代 np_i^0. 换言之，如果 $\hat{\boldsymbol{\theta}}$ 表示的是基于观察个数 $N_1,\cdots,N_k$ 的向量参数 $\boldsymbol{\theta}$ 的极大似然估计，那么统计量 θ 则定义如下：

$$Q=\sum_{i=1}^{k}\frac{[N_i-n\pi_i(\hat{\boldsymbol{\theta}})]^2}{n\pi_i(\hat{\boldsymbol{\theta}})}. \tag{9.2.4}$$

再者，基于该统计量 Q 来进行假设(9.2.3)的检验，即当 $Q\geqslant c$ 时，拒绝假设 H_0 是合理的，此处 c 是一个恰当的常数. 在1924年，费希尔证明了如果原假设 H_0 为真并且满足某些正则条件的话，那么当样本容量 $n\to\infty$ 时，$\boldsymbol{\theta}$ 的分布函数收敛于自由度为 $k-1-s$ 的 χ^2 分布.

因而，当样本容量很大而且原假设 H_0 为真时，$\boldsymbol{\theta}$ 的分布将近似地为一个 χ^2 分布. 为了决定自由度的值，因为在把 $i=1,\cdots,k$ 时的观测值 N_i 和期望值 $n\pi_i(\hat{\boldsymbol{\theta}})$ 作比较时，是要估计 s 个参数 $\theta_1,\cdots,\theta_s$，所以必须从9.1节中用到的 $k-1$ 中减去 s. 为了使该结果成立，满足以下正则条件是必要的：首先，向量 $\boldsymbol{\theta}$ 的极大似然估计 $\hat{\boldsymbol{\theta}}$ 必然出现在似然函数的对于每个参数 $\theta_1,\cdots,\theta_s$ 的偏导数都等于0的那个点. 其次这些偏导数必须满足极大似然估计的渐近性质. 544

作为使用(9.2.4)式中所定义的统计量 Q 的例题，我们将再次考虑在本节中曾描述过的关于两种类型基因的问题. 在第一种类型的一个问题中，$k=3$，要求检验原假设 H_0，即概率 p_1，p_2 和 p_3 可以表示成(9.2.1)中的形式，其对应的备择假设是 H_1，而 H_0 不是真的. 在该问题中，$s=1$. 因此，当 H_0 是真时，(9.2.4)式中所定义的统计量 Q 的分布将近似于自由度为3的 χ^2 分布.

在第二种类型的问题中，$k=6$，要求检验原假设 H_0，即概率 $p_1,\cdots,p_6$ 可以表示为(9.2.2)式的形式，对应的是备择假设 H_1 而 H_0 不是真的. 在该问题中，$s=2$. 因此，当 H_0 为真时，Q 的分布将近似于自由度为3的 χ^2 分布.

9.2.3 确定极大似然估计

当(9.2.3)式中的原假设 H_0 为真时，观测个数 $N_1,\cdots,N_k$ 的似然函数 $L(\boldsymbol{\theta})$ 是

$$L(\boldsymbol{\theta})=[\pi_1(\boldsymbol{\theta})]^{N_1}\cdots[\pi_k(\boldsymbol{\theta})]^{N_k}. \tag{9.2.5}$$

因而，

$$\ln L(\boldsymbol{\theta})=\sum_{i=1}^{k}N_i\ln\pi_i(\boldsymbol{\theta}). \tag{9.2.6}$$

极大似然估计 $\hat{\boldsymbol{\theta}}$ 是 $\ln L(\boldsymbol{\theta})$ 达到极大值时的 $\boldsymbol{\theta}$ 的值.

当 $k=3$，并且由 H_0 指定的概率 p_1，p_2 和 p_3 可以被表示成(9.2.1)式的形式时，那么

$$\ln L(\boldsymbol{\theta}) = N_1 \ln(\boldsymbol{\theta}^2) + N_2 \ln[2\boldsymbol{\theta}(1-\boldsymbol{\theta})] + N_3 \ln[(1-\boldsymbol{\theta})^2] \quad (9.2.7)$$
$$= (2N_1 + N_2)\ln \boldsymbol{\theta} + (2N_3 + N_2)\ln(1-\boldsymbol{\theta}) + N_1 \ln 2.$$

通过求导数可以得到，当 $\ln L(\boldsymbol{\theta})$ 是最大值时 $\boldsymbol{\theta}$ 的值为

$$\hat{\boldsymbol{\theta}} = \frac{2N_1 + N_2}{2(N_1 + N_2 + N_3)} = \frac{2N_1 + N_2}{2n}. \quad (9.2.8)$$

现在就可以由观察数 N_1，N_2 和 N_3 计算出(9.2.4)式中所定义的统计量 Q 的值. 如先前所提到的，当 H_0 为真且 n 很大时，Q 的分布将会近似为一个自由度为 1 的 χ^2 分布. 因此，相对于 Q 的观察值的尾部面积可以由该 χ^2 分布得到.

9.2.4 分布的正态性检验

现在考虑一个问题，从某个概率密度函数未知的连续分布中抽取一个随机样本 $X_1, \cdots, X_n$，要求检验分布服从正态分布这一原假设 H_0，其备择假设 H_1 为分布不服从正态分布. 可以用拟合优度的 χ^2 检验来处理该问题，如果将实数轴分成 k 个子区间并且对于随机样本落入第 i 个区间的观察值的个数 $N_i(i=1,\cdots,k)$ 进行计数.

545

如果 H_0 为真时，并且令 μ 和 σ^2 分别表示正态分布的未知的均值和方差，则参数向量 $\boldsymbol{\theta}$ 是二维向量 $\boldsymbol{\theta} = (\mu, \sigma^2)$. 观察值将落入第 i 个子区间的概率 $\pi_i(\boldsymbol{\theta})$ 或 $\pi_i(\mu, \sigma^2)$，即均值为 μ 和方差为 σ^2 的正态分布分配给该子区间的概率. 换言之，如果第 i 个子区间是从 a_i 到 b_i 的一个区间，那么

$$\pi_i(\mu, \sigma^2) = \int_{a_i}^{b_i} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx. \quad (9.2.9)$$

值得一提的是为了计算出由式(9.2.4)所定义的统计量 Q 的数值，极大似然估计量 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 必须通过不同子区间的观察值 $N_1, \cdots, N_n$ 来得到，而不能用 $X_1, \cdots, X_n$ 的观察值本身来得到. 换言之，$\hat{\mu}$ 和 $\hat{\sigma}^2$ 的值应该是使得似然函数取得最大值的 μ 和 σ^2 的值.

$$L(\mu, \sigma^2) = [\pi_1(\mu, \sigma^2)]^{N_1} \cdots [\pi_k(\mu, \sigma^2)]^{N_k}. \quad (9.2.10)$$

因为由式(9.2.9)中给出的函数 $\pi_i(\mu, \sigma^2)$ 的复杂性，为了确定能使 $L(\mu, \sigma^2)$ 的值达到最大的 μ 和 σ^2 的值需要一个很长的数值计算. 从另一方面来说，我们知道 μ 和 σ^2 的基于原始样本中 n 个观察值 $X_1, \cdots, X_n$ 的极大似然估计就是简单的样本均值 $\overline{X}_n$ 和样本方差 S_n^2/n. 此外，如果用使似然函数 $L(\mu, \sigma^2)$ 最大化的估计值来计算统计量 Q，那么我们知道当原假设 H_0 是正确时，Q 的分布将会趋于一个自由度为 $k-3$ 的 χ^2 分布. 另一方面，如果用基于原始样本观察值的极大似然估计来计算 Q，那么用 χ^2 分布来近似的分布就不太合适. 由于估计量 $\overline{X}_n$ 和 S_n^2/n 的简单性，我们将用这些估计值去计算 Q 的值，但是我们将描述它们的使用是如何修正 Q 的分布.

在 1954 年，察诺夫和莱曼得到如下的结果：如果用极大似然估计 $\overline{X}_n$ 和 S_n^2/n 来计算统计值 Q，并且如果原假设 H_0 是真的，那么当 $n \to \infty$，Q 的密度函数收敛于一

个介于自由度为 $k-3$ 的 χ^2 分布的密度函数和自由度为 $k-1$ 的 χ^2 分布的密度函数之间的一个密度函数. 由上可知，如果 Q 的值是通过该种简化的方法计算出来的，那么和 Q 值相对应的尾部面积事实上要比以自由度为 $k-3$ 的 χ^2 分布表中发现的尾部面积的数值来得大. 事实上，合适的尾部面积的值落在从自由度为 $k-3$ 的 χ^2 分布表中发现的尾部面积和从自由度为 $k-1$ 的 χ^2 分布表发现的更大的表中发现的尾部面积之间的某处. 因此，通过这种简化的办法来计算 Q 的值时，相应的尾部面积的值将以从 χ^2 分布表得到的这两个值为界.

例 9.2.1 滚珠轴承的失效时间 在例 9.1.2 中，我们检验了滚珠寿命的对数服从均值为 3.912 和方差为 0.25 的正态分布这一原假设，现在可以试着去检验，滚珠寿命的对数服从某个正态分布这一复合假设. 将实轴分为和例 9.1.2 中相同的子区间段，即分为 $(-\infty,3.575]$，$(3.575,3.912]$，$(3.912,4.249]$ 和 $(4.912,\infty)$. 落入每个区间观察值的个数仍为 3，4，8，8. 利用察诺夫和莱曼的结论，这样使得我们可以使用基于原始数据的极大似然估计. 由此可得 $\hat{\mu}=4.150$ 和 $\hat{\sigma}^2=0.2722$. 落在四个区间的概率分别为：

546

$$\pi_1(\hat{\mu},\hat{\sigma}^2)=\Phi\left(\frac{3.575-4.150}{(0.2722)^{1/2}}\right)=0.1350,$$

$$\pi_2(\hat{\mu},\hat{\sigma}^2)=\Phi\left(\frac{3.912-4.150}{(0.2722)^{1/2}}\right)-\Phi\left(\frac{3.575-4.150}{(0.2722)^{1/2}}\right)=0.1888,$$

$$\pi_3(\hat{\mu},\hat{\sigma}^2)=\Phi\left(\frac{4.249-4.150}{(0.2722)^{1/2}}\right)-\Phi\left(\frac{3.912-4.150}{(0.2722)^{1/2}}\right)=0.2511,$$

$$\pi_4(\hat{\mu},\hat{\sigma}^2)=1-\Phi\left(\frac{4.249-4.150}{(0.2722)^{1/2}}\right)=0.4251.$$

这使得 Q 值等于

$$Q=\frac{(3-23\times0.1350)^2}{23\times0.1350}+\frac{(4-23\times0.1888)^2}{23\times0.1888}+\frac{(8-23\times0.2511)^2}{23\times0.2511}+\frac{(8-23\times0.4251)^2}{23\times0.4251}=1.211.$$

与 1.211 相对应的尾部面积需要在以 $k-1=3$ 的自由度和 $k-3=1$ 的自由度的 χ^2 分布中算出. 对于自由度为 1 的 p-值是 0.2711，对于自由度为 3 的 p-值是 0.7504. 所以真正的 p-值落在区间 $[0.2711,0.7504]$ 中. 虽然该区间很大，但是它告诉我们在水平 α_0 下如果 $\alpha_0<0.2711$ 时，不要拒绝原假设. ◀

9.2.5 关于任意分布的复合假设的检验

刚刚所描述的过程可以被广泛地应用. 再次考虑这样一个问题，从某个概率密度函数未知的连续分布中抽取由 n 个观察值组成的一个随机样本. 假设现在希望检验该分布属于由 s - 维参数向量 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_s)$ 所标记的某个分布族这一原假设，其

备择假设 H_1 是该分布不属于那个特殊的分布族. 此外, 跟通常情况一样, 假定实数轴被分为 k 个子区域.

如果原假设 H_0 是真的, 而且向量 $\boldsymbol{\theta}$ 是通过式(9.2.5)中给定似然函数 $L(\boldsymbol{\theta})$ 取极大值来估计的, 那么统计量 Q 将会近似服从自由度为 $k-1-s$ 的 χ^2 分布. 然而, 如果 H_0 是真的, 并且用根据原来样本中 n 个观察值所得到的极大似然估计来计算统计量 Q, 那么 Q 的分布的一个合适的近似会落在分别以 $k-1-s$ 和 $k-1$ 为自由度的 χ^2 分布之间. 因此, 和该计算所得的 Q 值相对应, 从自由度为 $k-1-s$ 的 χ^2 分布表中所得的尾部面积将会是这个合适的尾部面积的下界, 而从自由度为 $k-1$ 的 χ^2 分布表所得的尾部面积将会是该合适尾部面积的一个上界.

547

刚才所描述的结论也同样适用于离散分布. 举个例子, 假设从一个可能取值为非负整数0,1,2,…的一个离散型分布中抽取一个有 n 个观察值的随机样本. 并设希望检验分布是一个泊松分布的原假设, 而其备择假设 H_1 是分布不是泊松分布. 最后, 假定非负整数0,1,2,…被分成 k 个组使得每个观察值落入这些组中的一组.

从6.4节的习题5可知, 如果 H_0 是真的, 那么基于原始样本中 n 个观察值的样本均值 $\overline{X}_n$ 就是泊松分布未知均值 $\boldsymbol{\theta}$ 的极大似然估计. 因此, 如果用估计量 $\boldsymbol{\theta}=\overline{X}_n$ 来计算式(9.2.4)中所定义的统计量 Q, 那么当 H_0 是真的, $\boldsymbol{\theta}$ 的近似分布落在自由度为 $k-2$ 和自由度为 $k-1$ 的 χ^2 分布之间.

9.2.6 小结

如果想要检验我们的数据是来自于一个参数族的分布这一复合假设, 必须要估计参数 $\boldsymbol{\theta}$. 首先, 我们把实数分成 k 个不相交的区间. 然后将数据简化成有多少观察值分别落入 k 个区间的计数 $N_1,\cdots,N_k$. 再构造一个似然函数 $L(\boldsymbol{\theta})=\prod_{i=1}^{k}\pi_i(\boldsymbol{\theta})^{N_i}$, 其中 $\pi_i(\boldsymbol{\theta})$ 是一个观察值落入第 i 个区间中的概率. 估计 $\boldsymbol{\theta}$ 的值为使 $L(\boldsymbol{\theta})$ 达到最大的值 $\hat{\boldsymbol{\theta}}$. 然后计算具有 $\sum$(观察值 - 期望值)^2/期望值形式的检验统计量 $Q=\sum_{i=1}^{k}[N_i-n\pi_i(\hat{\boldsymbol{\theta}})]^2/[n\pi_i(\hat{\boldsymbol{\theta}})]$. 为了在 α_0 水平下检验原假设, 要将 $\boldsymbol{\theta}$ 的值与自由度为 $k-1-s$ 的 χ^2 分布的 $1-\alpha_0$ 分位点的值作比较, 其中 s 是 $\boldsymbol{\theta}$ 的维数. 另一个方法是, 基于原始观察值可以找到通常的极大似然估计 $\hat{\boldsymbol{\theta}}$. 在这种情况下, 需要将 Q 与自由度为 $k-1-s$ 的 χ^2 分布的 $1-\alpha_0$ 分位数的值和自由度为 $k-1$ 的 χ^2 分布的 $1-\alpha_0$ 分位数之间的一个数作比较.

习题

1. 表9-4中的41个数字是美国41个城市中所测量出的空气中二氧化硫的平均含量(微克每立方米). 数据由 Sokal and Rohlf(1981)的第619~620页给出.
   a. 检验"数据来自正态分布"的原假设.
   b. 检验"数据来自对数正态分布"的原假设.

表 9-4 美国 41 所城市空气中二氧化硫的含量

10	13	12	17	56	36	29	14	10	24	110	28	17	8
30	9	47	35	29	14	56	14	11	46	11	23	65	26
69	61	94	10	18	9	10	28	31	26	29	31	16	

2. 在第 5 届曲棍球比赛中，在赛场，随机挑选了 200 人，并询问他们在此之前的 4 场比赛中参加了几场. 结果在表 9-5 中给出. 检验"将 200 个观察值看作是来自于二项分布的一个随机样本"这一原假设，即存在一个数 θ $(0<\theta<1)$，使得可以将概率表示为：
$p_0=(1-\theta)^4, p_1=4\theta(1-\theta)^3, p_2=6\theta^2(1-\theta)^2, p_3=4\theta^3(1-\theta), p_4=\theta^4$.

548

表 9-5 习题 2 的数据

以前参加的比赛数	0	1	2	3	4
人数	33	67	66	15	19

3. 考虑一个遗传问题，在某个总体中每个个体都必须有六种基因中的一种. 检验原假设 H_0：即六种基因的概率能被表示成式(9.2.2)的形式.
   a. 假定在一个有 n 个个体的随机样本中，具有六种基因的个体人数的观察值分别为 $N_1,\cdots,N_6$，求出当原假设 H_0 为真时，参数 θ_1，θ_2 的极大似然估计.
   b. 假定在一个有 150 个个体的随机样本中，观察值如下：
   $$N_1=2, N_2=36, N_3=14, N_4=36, N_5=20, N_6=42.$$
   求出 θ 的值及其相应的尾部面积.
4. 再次考虑 9.1 节中习题 8 中由 500 个男人身高的数据组成的样本. 假定在该习题中的数据被分割成区间前，可以发现对于原始样本中 500 个观察到的身高，其样本均值为 $\overline{X}_n=67.6$，样本方差 $S_n^2/n=1.00$. 检验假设"这些被观察到的身高所组成的随机样本来自于正态分布".
5. 在一个大城市，随机选取 200 个人，并询问每个人在本周购买了多少张彩票. 结果如表 9-6中所给. 假定在购买了 5 张或更多彩票的 7 个人中，有 3 个人正好买了 5 张，2 个人买了 6 张，一个人买了 7 张，还有一个人买了 10 张. 检验假设"由这 200 个观察值所组成的随机样本来自于泊松分布".

表 9-6 习题 5 的数据

先前买的彩票数	0	1	2	3	4	5 或更多
人数	52	60	55	18	8	7

6. Rutherfor and Geiger (1910) 收集了在 2608 个不相交的时间段内，镭元素发射 α 粒子的个数，每个时间段的长度为 7.5 秒. 结果如表 9-7 所示. 检验假设"这 2608 个观察值组成的随机样本来自于泊松分布".

表 9-7 习题 6 来自于 Rutherford and Gerger (1910) 的数据

发射的粒子数	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15 或更多	总计
时间段的个数	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1	0	2608

549

7. 检测假设“由表9-8中的50个观察值所组成的随机样本来自于正态分布”.

表9-8　习题7的数据

9.69	2.78	5.21	4.07	9.81	0.52	10.22	9.09	3.03	16.80
8.93	7.47	2.62	5.15	5.61	6.80	5.05	9.20	5.29	8.07
7.61	8.46	0.22	8.32	13.98	2.90	6.06	7.82	8.68	0.66
8.12	7.89	-0.59	6.01	10.22	2.06	14.51	8.67	11.81	4.01
-2.74	5.93	8.77	0.68	7.89	11.15	13.05	7.52	7.80	8.64

8. 检验假设“表9-9中50个观察值所组成的随机样本来自于指数分布”.

表9-9　习题8的数据

0.91	0.90	0.16	1.62	1.28	1.16	3.65	0.61	1.75	0.75
1.22	0.86	2.02	0.56	0.06	0.22	3.41	0.31	0.89	0.05
1.28	1.45	1.59	0.53	0.19	0.91	0.07	0.22	0.79	1.53
0.22	1.22	1.73	0.50	0.29	0.04	0.51	0.37	1.28	1.86
2.33	0.55	0.49	0.24	0.74	1.41	1.27	0.06	0.57	1.28

9.3 列联表

当样本中的每一个观察值都是一个二元离散型随机向量(一对离散型随机变量)，那么，有一个简单的方法来检验这两个随机变量是独立的这一假设. 该检验就是本章前面用过的χ^2检验的另一种形式.

9.3.1 列联表中的独立性

假设从一个大学的全部入学新生中随机地选择200名学生，并且样本中的每个学生都按照其入学时课程及其在即将到来的选举中对候选人A和B的偏好来分类. 假设结果如表9-10所示.

表9-10　按照课程和对候选人的偏爱对学生进行分类

课　程	偏爱的候选人			总　数
	A	B	未决定	
工程和科学	24	23	12	59
人类和自然科学	24	14	10	48
艺术	17	8	13	38
工商管理	27	19	9	55
总数	92	64	44	200

一个表格中每个观察值都被分成两种或更多种的因素时被称为列联表. 在表9-10中，对于每个学生只有两种分类，也就是学生入学时的课程和他们喜欢的候选人. 称这种表格为一个两因素的列联表.

当一个统计员分析一个列联表时，她经常对于检验不同分类是否独立这一假设

感兴趣．因此，对于表9-10，她可能对检验假设“每个学生人学时课程和他们喜欢的候选人是否是独立的变量”感兴趣．对于这个问题，可以如下准确地描述独立的假设．假设在大学的所有入学学生中，随机选择一个学生．那么对于每个 i 和 j，这样一个随机选择的学生偏爱候选人 j 的概率和他在申请课程 i 条件下偏爱候选人 j 的条件概率是相同的．

550

通常来说，应该将一个两因素的列联表分为 R 行和 C 列．对于 $i=1,\cdots,R$ 和 $j=1,\cdots,C$，令 p_{ij} 表示从一个给定总体中随机抽取的个体会被分在表中第 i 行和第 j 列的概率．此外，可以令 p_{i+} 表示该个体会被分类在表格的第 i 行的边际概率，且令 p_{+j} 表示该个体会被分类在表格的第 j 列的边际概率．因此，

$$p_{i+}=\sum_{i=1}^{C}p_{ij}\text{和}\,p_{+j}=\sum_{i=1}^{R}p_{ij}.$$

此外，由于表中所有元素的可能值的总数等于1，我们有

$$\sum_{i=1}^{R}\sum_{j=1}^{C}p_{ij}=\sum_{i=1}^{R}p_{i+}=\sum_{j=1}^{C}p_{+j}=1.$$

现在假设从给定的总体中抽取一个有 n 个个体的随机样本，对于 $i=1,\cdots,R$ 和 $j=1,\cdots,C$，我们令 N_{ij} 表示被分类到表中第 i 行第 j 列的个体数．另外，令 N_{i+} 表示在第 i 行的个数总和，而 N_{+j} 表示在第 j 列的个数总和．因此

$$N_{i+}=\sum_{j=1}^{C}N_{ij}\text{和}\,N_{+j}=\sum_{i=1}^{R}N_{ij}.\tag{9.3.1}$$

还有，

551

$$\sum_{i=1}^{R}\sum_{j=1}^{C}N_{ij}=\sum_{i=1}^{R}N_{i+}=\sum_{j=1}^{C}N_{+j}=n.\tag{9.3.2}$$

在这些观察值的基础上，要检验下面这些假设：

$$\begin{aligned}&H_0: p_{ij}=p_{i+}p_{+j},\text{对于}\ i=1,\cdots,R\ \text{和}\ j=1,\cdots,C,\\&H_1:\text{原假设}\ H_0\ \text{不真}.\end{aligned}\tag{9.3.3}$$

9.3.2 独立性的 χ^2 检验

在9.2节中所描述的 χ^2 检验可以被应用到检验假设(9.3.3)的问题中．从总体中抽取的样本必须来自列联表中 RC 个单元中的一个．在原假设 H_0 条件下这些单元的未知概率 p_{ij} 可以被表示为未知参数 p_{i+} 和 p_{+j} 的函数．由于 $\sum_{i=1}^{R}p_{i+}=1$ 和 $\sum_{j=1}^{C}p_{+j}=1$，所以当 H_0 为真时，要被估计的未知参数个数为 $(R-1)+(C-1)$，或者是为 $R+C-2$．

对 $i=1,\cdots,R$ 和 $j=1,\cdots,C$，当 H_0 为真时，令 $\hat{E}_{ij}$ 表示会被分到表中第 i 行和第 j 列中观察值的期望值的极大似然估计．在此问题中，由公式(9.2.4)所定义的统计量 Q 将具有以下的形式

$$Q=\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{(N_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}}.\tag{9.3.4}$$

此外，因为列联表中包含了 RC 个单元，而且由于当 H_0 为真时，要估计值的参数个数为 $R+C-2$ 个，所以可知当 H_0 为真且 $n\to\infty$ 时，统计量 Q 的密度函数收敛于自由度为 $RC-1-(R+C-2)=(R-1)(C-1)$ 的 χ^2 分布的密度函数.

接着，要考虑估计量 $\hat{E}_{ij}$ 的形式，在第 i 行和第 j 列的观察值的期望数可以被简记为 np_{ij}. 当 H_0 为真时，$p_{ij}=p_{i+}p_{+j}$. 因此，如果 $\hat{p}_{i+}$ 和 $\hat{p}_{+j}$ 分别表示 p_{i+} 和 p_{+j} 的极大似然估计的估计值，那么它必须满足 $\hat{E}_{ij}=n\hat{p}_{i+}\hat{p}_{+j}$. 再次，因为 p_{i+} 是一个观察值会被分在第 i 行的概率，而 $\hat{p}_{i+}$ 是在样本中观察值会被分在第 i 行的比例，即 $\hat{p}_{i+}=N_{i+}/n$. 同理，$\hat{p}_{+j}=N_{+j}/n$. 由此可得

$$\hat{E}_{ij}=n\left(\frac{N_{i+}}{n}\right)\left(\frac{N_{+j}}{n}\right)=\frac{N_{i+}N_{+j}}{n}. \tag{9.3.5}$$

如果把 $\hat{E}_{ij}$ 的这个值代入式(9.3.4)中，就可从 N_{ij} 的观察值中计算出 Q 的值. 如果 $Q>c$，这里 c 是一个适当选取的常数，就拒绝原假设. 当 H_0 为真且样本容量 n 较大时，Q 的分布将会近似是一个自由度为 $(R-1)(C-1)$ 的 χ^2 分布.

552

例 9.3.1　学生的分类　如果我们希望用表 9-10 的数据对假设(9.3.3)作检验. 用表中所给定的总数，可以发现 $N_{1+}=59, N_{2+}=48, N_{3+}=38$ 和 $N_{4+}=55$；以及 $N_{+1}=92, N_{+2}=64$ 和 $N_{+3}=44$. 因为 $n=200$，由式(9.3.5)可以算出如表 9-11 所示的 4×3 的 $\hat{E}_{ij}$ 值的表.

表 9-11　例 9.1.3 中期望的单元数

课　程	偏爱的候选人			总　数
	A	B	未决定	
工程与科学	27.14	18.88	12.98	59
人类和社会科学	22.08	15.36	10.56	48
艺术	17.48	12.16	8.36	38
工商管理	25.30	17.60	12.10	55
总数	92	64	44	200

现在表 9-10 中所给的 N_{ij} 的值可以与表 9-11 中 $\hat{E}_{ij}$ 的值作比较，由(9.3.4)式中所定义的 Q 的值结果为 6.68. 因为 $R=4$，$C=3$，所以可以从自由度为 $(R-1)(C-1)=6$ 的 χ^2 分布的表中找到相应的尾部面积. 它的值大于 0.3. 因此，如果 $\alpha_0\geqslant0.3$，在检验水平 α_0 下，只能拒绝原假设. ◀

例 9.3.2　蒙大拿州的民意调查　在例 9.1.1 中，我们做了蒙大拿州居民个人财政水平的民意调查. 另一个调查者被问及的问题是收入范围. 表 9-12 给出了一张上述两个问题回答的交叉表. 可以用 χ^2 检验来检验原假设“关于个人财政水平的观点和收入是独立的”. 在原假设下，表 9-13 给出了表 9-12 中每个单元的期望值. 现在可以根据自由度 $(3-1)\times(3-1)=4$ 来计算出检验统计量 $Q=5.210$. 与该 Q 值相关联的 p-值为 0.266，所以在一个大于 0.266 的水平 α_0 下拒绝原假设. ◀

表 9-12 蒙大拿州民意调查对两个问题的反映

收入范围	个人经济状况			总 数
	不好	一般	较好	
\$20 000 以下	20	15	12	47
\$20 000 - \$35 000	24	27	32	83
\$35 000 以上	14	22	23	59
总数	58	64	67	189

553

表 9-13 在独立的假设情况下表 9-12 每个单元的期望数

收入范围	个人经济状况			总 数
	不好	一般	较好	
\$20 000 以下	14.42	15.92	16.66	47
\$20 000 - \$35 000	25.47	28.11	29.42	83
\$35 000 以上	18.11	19.98	20.92	59
总数	58	64	67	189

9.3.3 小结

我们学习了如何检验“基于一个 n 对随机样本的两个离散型随机变量是独立的”这一原假设. 首先建立一个关于每对可能观察值的频数的列联表. 然后估计这两个随机变量的边际分布. 在随机变量是相互独立的原假设下，第一个变量 i 值和第二个变量 j 值的期望频数是 n 乘以两个估计的边际概率. 然后通过把列联表中所有单元的(观察值 - 期望值)^2/期望值相加来建立 χ^2 统计量 Q. χ^2 分布的自由度为$(R-1)(C-1)$，其中 R 等于表中的行数，C 等于表中的列数.

习题

1. Chase and Dummer(1992)年在研究了曼哈顿学龄儿童的想法. 他们询问儿童对于他们来说下列哪个是最重要的：成绩好，运动能力强和受欢迎. 同时也收集了关于每个儿童的其他信息，表 9-14 表示了 478 名儿童按其性别和不同的答案归类的结果. 检验原假设是“一个儿童对调查问题的答案和他或她的性别是无关的”.

表 9-14 由 Chase 和 Dummer (1992) 所提供的习题 1 中的数据

	成绩好	运动能力强	受欢迎
男孩	117	60	50
女孩	130	30	91

2. 证明由式(9.3.4)中所定义的统计量 Q 可以重写如下：

$$Q=\left(\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{N_{ij}^{2}}{\hat{E}_{ij}}\right)-n.$$

3. 证明：如果 $C=2$，由式(9.3.4)中所定义的统计量 Q 可以重写成：

$$Q = \frac{n}{N_{+2}}\left(\sum_{i=1}^{R}\frac{N_{i1}^2}{\hat{E}_{i1}} - N_{+1}\right).$$

4. 假设做一个试验来考查一个男人的年龄和他蓄须(留胡子)是否有关. 假定从大于等于18岁的男人中随机挑选100名；每一个男人按他是否在18到30岁之间以及他是否蓄须来归类. 在表9-15中给出了观测值. 检验假设：男人的年龄和他蓄不蓄须之间是没有关系的.

554

表9-15　习题4中的数据

	蓄须	不蓄须
18~30岁	12	28
超过30岁	8	52

5. 假设从一个很大的总体中随机地挑选300人，样本中的每个人按其血型 O，A，B 或者是 AB 和 Rh 的正负性来归类. 表9-16给出了观察值. 检验假设：血型的两种分类之间是独立的.

表9-16　习题5中的数据

	O	A	B	AB
Rh 正值	82	89	54	19
Rh 负值	13	27	7	9

6. 假设一个商店出售同一类型的两种不同品牌 A 和 B 的早餐麦片. 假如在一周时间内，商店记录卖掉的每一袋此类型的麦片是 A 品牌还是 B 品牌，同时还记录顾客是男人还是女人(如果是一个小孩买的,或是一个男人和一个女人一起买的,那就不记录在内). 假如卖了44袋，结果如表9-17所示. 检验假设：顾客的性别和购买的品牌之间是相互独立的.

表9-17　习题6中的数据

	品牌 A	品牌 B
男人	9	6
女人	13	16

7. 考虑一个两因素3行3列的列联表. 假如对 $i=1, 2, 3$ 和 $j=1, 2, 3$，从一个给定的总体中随机抽取一个个体会被归类到表9-18中第 i 行和第 j 列的概率 p_{ij}.

表9-18　习题7中的数据

0.15	0.09	0.06
0.15	0.09	0.06
0.20	0.12	0.08

a. 通过证明 p_{ij} 的值满足式(9.3.3)的原假设来证明该表的行和列是相互独立的.

b. 用一个均匀的伪随机数产生器，从一个给定的总体中生成一个有300个观察值的随机样本. 选择300个在0到1之间的伪随机数且操作如下：因为 $p_{11}=0.15$，如果 $x<0.15$ 就把伪随机数 x 归在第一个单元里. 因为 $p_{11}+p_{12}=0.24$，那么如果 $0.15\leqslant x<0.24$ 就

把伪随机数 x 归在第二个单元里. 对于九个单元都使用该方法. 例如, 除了 p_{33} 以外的所有概率的总和是 0.92, 所以如果 $x \geqslant 0.92$ 就把伪随机数 x 归在表中右下角的单元里.

c. 考虑由 b. 部分产生的 3×3 的表中的观察值 N_{ij}. 假如概率 p_{ij} 未知, 那么检验假设 (9.3.3).

8. 如果一个班级的所有学生独立地完成习题 7, 而且每个人用不同的伪随机数, 那么不同的学生得到的统计量 Q 的不同的值就会形成一个来自于自由度为 4 的 χ^2 分布的随机样本. 如果你能得到班级中所有学生的 Q 值, 试检验假设: 这些值形成了这样的一个随机样本.

9. 考虑一个三因素容量为 $R \times C \times T$ 的列联表. 对 $i=1,\cdots,R, j=1,\cdots,C$ 和 $k=1,\cdots,T$, 令 p_{ijk} 表示从一个给定的总体抽取一个个体会落入表中第 (i,j,k) 单元的概率. 设

$$p_{i++} = \sum_{j=1}^{C}\sum_{k=1}^{T} p_{ijk}, p_{+j+} = \sum_{i=1}^{R}\sum_{k=1}^{T} p_{ijk}, p_{++k} = \sum_{i=1}^{R}\sum_{j=1}^{C} p_{ijk}.$$

基于从给定的总体中抽取的由 n 个样本观察值所组成的一个随机样本, 建立如下假设的一个检验: $H_0: p_{ijk} = p_{i++}p_{+j+}p_{++k}$, 对于 i,j,k 的所有值都成立,

H_1: 假设 H_0 不正确.

555

10. 重新考虑习题 9 的条件. 对 $i=1,\cdots,R, j=1,\cdots,C$, 令 $p_{ij+} = \sum_{i=1}^{T} p_{ijk}$. 基于从给定的总体中抽取的 n 个样本观察值所组成的一个随机样本, 建立如下假设的一个检验:

$H_0: p_{ijk} = p_{ij+}p_{++k}$, 对于每一个 i,j 和 k 的值都成立,

556

H_1: 假设 H_0 不正确.

9.4 符号检验与秩检验

在本节中, 我们描述了一些常用的关于分布的中位数或者关于两种分布间差别的假设的非参数检验.

587

9.4.1 单样本过程

假设 $X_1,\cdots,X_n$ 是来自于一个未知分布的随机样本. 在第 8 章中, 我们考虑了未知分布的形式是已知的, 但是某些特定的参数仍然是未知的情形. 例如, 某一分布可能是均值和/或方差均未知的正态分布, 现在, 我们只假定分布是连续的. 由于不假定此数据的分布有一个均值, 所以不能对该分布的均值进行假设检验. 然而每个连续分布都有一个中位数 μ 满足 $\Pr(X_i \leqslant \mu) = 0.5$. 对于一般分布中位数是位置的一个常用的度量, 现在我们给出对于下列形式的假设的检验程序:

$$H_0: \mu \leqslant \mu_0, \qquad (9.4.1)$$
$$H_1: \mu > \mu_0.$$

该检验是基于下面这个简单事实的: $\mu \leqslant \mu_0$ 当且仅当 $\Pr(X_i \leqslant \mu_0) \geqslant 0.5$ 时, 对于 $i =$

$1,\cdots,n$,当 $X_i\leqslant\mu_0$ 时，令 $Y_i=1$，其他情况下令 $Y_i=0$. 定义 $p=\Pr(Y_i=1)$. 进而检验是否有 $\mu\leqslant\mu_0$ 等价于检验是否 $p\geqslant0.5$. 因为 $X_1,\cdots,X_n$ 是相互独立的，同样 $Y_1,\cdots,Y_n$ 也是. 这使得 $Y_1,\cdots,Y_n$ 是一个服从参数为 p 的伯努利分布的随机样本. 我们已经知道如何来检验 $p\geqslant0.5$ 这一原假设（参见例 8.1.5）. 计算 $W=Y_1+\cdots+Y_n$,并且如果 W 太小就拒绝原假设. 为了在显著性水平 α_0 下进行检验，选择 c 使得

$$\sum_{w=0}^{c}\binom{n}{w}\left(\frac{1}{2}\right)^n\leqslant\alpha_0<\sum_{w=0}^{c+1}\binom{n}{w}\left(\frac{1}{2}\right)^n.$$

如果 $W\leqslant c$，就拒绝原假设 H_0.

刚才所描述的检验称为符号检验，这是因为它是基于 $X_i-\mu_0$ 为负的观察值个数的基础上. 如果想检验假设

$$H_0:\mu=\mu_0,$$
$$H_1:\mu\neq\mu_0.$$

那么可以建立一个类似的检验. 再次令 $p=\Pr(X_i\leqslant\mu_0)$. 现在的原假设 H_0 等价于 $p=0.5$. 为了在显著性水平 α_0 下进行检验，要选择 c 使得

$$\sum_{w=0}^{c}\binom{n}{w}\left(\frac{1}{2}\right)^n\leqslant\frac{\alpha_0}{2}<\sum_{w=0}^{c+1}\binom{n}{w}\left(\frac{1}{2}\right)^n.$$

当 $W\leqslant c$ 或 $W\geqslant n-c$ 时，就拒绝 H_0. 我们用对称的拒绝域是因为参数为 n 和 1/2 的二项分布关于 $n/2$ 是对称的.

588

例 9.4.1 热狗的卡路里数 考虑 7.5 节习题 7 中 $n=20$ 时牛肉热狗的卡路里数. 假设我们对于检验原假设"卡路里含量的中位数是 150"，而其备择假设为"卡路里含量的中位数不是 150". 因为 20 只热狗中有 9 只热狗的卡路里含量比 150 要小，可知 $W=9$. 该观察值双边的 p-值是 0.8238，所以除非 $\alpha_0\geqslant0.8238$，否则不能在显著性水平 α_0 时拒绝原假设. ◀

对于每个 $\Pr(X_i\leqslant\mu_0)$ 的值很容易计算出符号检验的功效函数. 例如，在假设(9.4.1)的单边假设检验中，其功效是

$$\Pr(W\leqslant c)=\sum_{w=0}^{c}\binom{n}{w}p^w(1-p)^{n-w}.$$

9.4.2 比较两个分布

接下来，要考虑有 m 个观察值 $X_1,\cdots,X_m$ 的一个随机样本来自于一个分布函数 $F(x)$ 未知的连续分布，并且有 n 个观察值 $Y_1,\cdots,Y_n$ 的独立随机样本来自于另一个分布函数 $G(x)$ 也未知的连续分布. 要检验假设

$$H_0:F=G,$$
$$H_1:F\neq G. \tag{9.4.2}$$

检验假设(9.4.2)的一个方法是对两样本的 Kolmogorov-Smirnov 检验（本书未涉

及). 此外，如果我们愿意假定两样本事实上是取自有相同的未知方差的正态分布，那么检验假设(9.4.2)和检验两正态分布是否有同样的均值的检验是相同的. 因此，在此假定下，我们可以用8.4节所描述的两样本 t 检验.

在本部分我们将介绍检验假设(9.4.2)的另一个程序. 该过程大约在1940年由F. Wilcoxon和H. B. Mann与D. R. Whitney分别提出的，被称作Wilcoxon-Mann-Whitney的秩检验.

Wilcoxon-Mann-Whitney 秩检验

在本过程中，首先将两个样本中的 $m+n$ 个观察值按照它们在将两样本中出现的从最小值到最大值排列成单个序列. 因为所有的观察值均来自于连续分布，所以假定 $m+n$ 个观察值中没有两个的值是相等的. 因此，可以得到一个关于这 $m+n$ 个值的总体次序关系. 在总次序中的每个观察值按照它在次序中的位置都分别赋予了从1到 $m+n$ 的序号.

Wilcoxon-Mann-Whitney 秩检验的特点在于：如果假定原假设 H_0 为真，两个样本事实上来自于同一分布，那么观察值 $X_1,\cdots,X_m$ 在 $m+n$ 个观察值的排序中会有分散开的倾向，而不是集中在较小值或较大值附近. 事实上，若 H_0 为真，分配给这 m 个观察值 $X_1,\cdots,X_m$ 的秩就如同随机地从一个放有 $m+n$ 个等级 $1,2,\cdots,m+n$ 的盒子中无放回地抽取 m 个等级的一个随机样本.

589

令 S 表示为分配给 m 个观察值 $X_1,\cdots,X_m$ 的秩和. 因为秩 $1,2,\cdots,m+n$ 的平均值为 $(1/2)(m+n+1)$，由刚才所给的讨论中可知：当 H_0 为真时，

$$\mathrm{E}(S)=\frac{m(m+n+1)}{2}. \tag{9.4.3}$$

还可以证明当 H_0 为真时，

$$\mathrm{Var}(S)=\frac{mn(m+n+1)}{12}. \tag{9.4.4}$$

此外，当样本容量 m，n 都很大且 H_0 为真时，S 的分布将近似于均值和方差分别由(9.4.3)式和(9.4.4)式给出的正态分布. 如果 S 的值偏离(9.4.3)式中所给定的均值很大，Wilcoxon-Mann-Whitney 秩检验就拒绝 H_0. 即，选取一个恰当的常数 c，如果 $|S-(1/2)m(m+n+1)|\geqslant c$,检验就明确地拒绝原假设 H_0. 特别地，若采用 S 的近似正态分布，则常数 $c=[\mathrm{Var}(S)]^{1/2}\Phi^{-1}(1-\alpha_0/2)$ 使得 Wilcoxon 检验的显著性水平为 α_0.

例 9.4.2 比较下列铜矿石 再次考虑例8.4.2中的铜矿石的比较. 假定在第一个样本中 $m=8$，所测得数据为：

2.183, 2.431, 2.556, 2.629, 2.641, 2.715, 2.805, 2.840,

而在第二个样本中 $n=10$ 时所测得数据为：

2.120, 2.153, 2.213, 2.240, 2.245, 2.266, 2.281, 2.336, 2.558, 2.587.

两个样本中18个值按从小到大排列于表9-19中. 第一个样本中每个观察值用符号 x

来表示，第二个样本中每个观察值用符号 y 表示．第一个样本中10个观察值的秩和 S 为104.

假定用正态分布来近似．那么当 H_0 为真，S 近似地有均值为76、方差为126.67的正态分布．因此 S 的标准差为 $(126.67)^{1/2}=11.25$. 所以，当 H_0 为真时，随机变量 $Z=(S-76)/(11.25)$ 将近似地服从标准正态分布．由于在这个例题中 $S=104$，所以可知 $Z=2.49$. 和该 Z 值相对应的 p-值为0.0128．因此，在每一个 $\alpha_0 \geqslant 0.0128$ 的显著性水平下都将要拒绝原假设 H_0. ◀

对于 m，n 小的数值采用正态分布来近似 S 的分布是不恰当的．在许多已出版的统计表的收集中给出了小样本量的 S 的精确分布表．许多统计软件包也可以计算概率分布函数和 S 的精确分布的分位数.

注：成对数据的检验．在习题1和习题15中，有成对数据的符号检验与秩检验的补充说明.

590

表9-19　例9.4.2中的分类数据

序号	观察到的数值	样本	序号	观察到的数值	样本
1	2.120	y	10	2.431	x
2	2.153	y	11	2.556	x
3	2.183	x	12	2.558	y
4	2.213	y	13	2.587	y
5	2.240	y	14	2.629	x
6	2.245	y	15	2.641	x
7	2.266	y	16	2.715	x
8	2.281	y	17	2.805	x
9	2.336	y	18	2.840	x

9.4.3　同分值

Wilcoxon-Mann-Whitney 符号秩检验的理论是建立在所有的观察值 X_i，Y_j 都是不同的假设上的．然而，由于在实际的试验中只能以有限的精确性进行测量，因此确实存在同一个观察值不止出现一次的情况．例如：假设要做一个 Wilcoxon-Mann-Whitney 秩检验，发现有一对或多对 (i,j) 使得 $X_i=Y_j$．在此情况下，就要进行两次秩检验．在第一次检验中，对于每一对 $X_i=Y_j$，应当假定每一个 $X_i<Y_j$．在第二次检验中，假定 $X_i>Y_j$．如果两个检验的尾部面积大致相同，那么同分值在数据中是一个相对不太重要的部分．另一方面，如果尾部面积差别很大，那么同分值对于所要做出的推断影响很大．在这种情况下数据有可能是缺乏说服力的.

例9.4.3　钙的补充和血压　回顾8.4节习题10中的数据．在两个样本中都有 -5 和 -3 这两个观察值．首先，把较小的秩分配给接受钙补充的小组的观察值（X_i），

然后，把余下较小的秩分配给服食安慰剂小组的观察值(Y_j). 例如，在复合样本中，-3 是第5，第6，第7小的值. 在第一次检验中，把秩5分配给 $X_i=-3$，把秩6、7分配给两个 $Y_j=-3$，在第二次检验中，把秩5、6分配给 $Y_j=-3$. 对于第一个检验，X 的秩之和为123，在第二个检验中，X 的秩之和为126. 在此问题中，$m=10$，$n=11$，所以当原假设为真时，S 的均值和方差分别为110和201.7. 对应这两个值的双边尾部面积分别为0.36，0.26. 除非 $\alpha_0 \geq 0.26$，否则这两种都不会导致在水平 α_0 下拒绝原假设. ◀ 591

处理同分值也有其他的合理的方法. 当有两个或更多值相等时，一个简单的方法是把连续的秩分配给这些值，然后将这些秩的平均值分配给同分值，当用这个方法时，由于同分值的原因，Var(S)的值必须要修正.

9.4.4 扩展研究：Wilcoxon-Mann-Whitney 秩检验的功效

当 X 的秩总和 S 太大或太小时，Wilcoxon-Mann-Whitney 秩检验就拒绝认为"两个分布是相同的"这一原假设. 认为最重要的备择假设是那些"X_i 的值趋向于比 Y_j 的值大"或"X_i 的值趋向于比 Y_j 的值小"的假设是很敏感的事. 然而还有其他当 $F \neq G$ 的情况，但 S 趋向于式(9.4.3)的均值. 例如：假设所有的 $X_1,\cdots,X_m$ 都服从于区间[0,1]上的均匀分布以及 $Y_1,\cdots,Y_n$ 有如下的概率密度函数：

$$g(y)=\begin{cases}0.5, & \text{当 } -1<y<0 \text{ 或 } 1<y<2, \\ 0, & \text{其他.}\end{cases}$$

那么就不难证明 E(S)和等式(9.4.3)相同且 $\mathrm{Var}(S)=m^2n/4$. 在这种情况下，检验的功效（拒绝 H_0 的概率）也不会明显大于显著性水平 α_0. 事实上，如果关注该类型的备择假设，那么只要 X 的秩太紧密地聚集在一起而不管他们是大还是小，都希望拒绝原假设.

Wilcoxon-Mann-Whitney 秩检验被设计为当 F 和 G 之间有特殊关系时具有较大的功效. 令 X 表示具有分布函数 F 的随机变量，且令 Y 表示具有分布函数 G 的随机变量. 令 F^{-1} 和 G^{-1} 分别表示对应的分位函数. 如果对于所有的 $0<p<1$ 有 $F^{-1}(p) \geq G^{-1}(p)$. 则我们说 F 随机地大于 G，或等价地 X 随机地大于 Y，即，每一个 X 的分位数至少和相应的 Y 的分位数一样大. 显而易见，如果 X_i 随机地大于 Y_j，那么在混合样本中 X_i 的秩至少和 Y_j 的秩一样大. 这就使得 S 的值大的可能性比小的可能性更大. 类似地，如果 Y_j 是随机地大于 X_i，S 将趋向于小的值.

当 X_i 或 Y_j 都不是随机地比另一个大，那么就很难给 S 的分布下一个一般的论断. 对于大容量样本，即使当 $F \neq G$ 时，对 S 的分布正态的近似依旧是成立的. 然而 S 的均值和方差依赖于两个分布函数 F 和 G. 例如，用习题11中的结论可以证明

$$\mathrm{E}(S)=nm\,\Pr(X_1 \geq Y_1)+\frac{m(m+1)}{2}. \tag{9.4.5}$$

592

用同样的方法，还可以证明：

$$\operatorname{Var}(S)=nm(\Pr(X_1\geqslant Y_1)+(1+m+n)\Pr(X_1\geqslant Y_1)^2+ \\ (m-1)\Pr(X_1\geqslant Y_1,X_1\geqslant Y_2)+(n-1)\Pr(X_1\geqslant Y_1,X_2\geqslant Y_2)). \tag{9.4.6}$$

原则上，F 和 G 每一个特定的选择都能计算出这些概率，可以用模拟的方法(见第11章)来近似得到所需要的概率. 在计算或近似得到这些概率后，就可以近似得到如下所示的显著性水平为 α_0 的 Wilcoxon-Mann-Whitney 秩检验的功效. 首先，回忆如果 $S\leqslant c_1$ 或 $S\geqslant c_2$ 时，拒绝 $F=G$ 的原假设. 其中

$$c_1=\frac{m(m+n+1)}{2}-\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\left[\frac{mn(m+n+1)}{12}\right]^{1/2},$$

$$c_2=\frac{m(m+n+1)}{2}+\Phi^{-1}\left(1-\frac{\alpha_0}{2}\right)\left[\frac{mn(m+n+1)}{12}\right]^{1/2}.$$

那么这个检验的功效是

$$\Phi\left(\frac{c_1-\mathrm{E}(S)}{\operatorname{Var}(S)^{1/2}}\right)+1-\Phi\left(\frac{c_2-\mathrm{E}(S)}{\operatorname{Var}(S)^{1/2}}\right),$$

其中 $\mathrm{E}(S)$ 和 $\operatorname{Var}(S)$ 分别由等式(9.4.5)和(9.4.6)给定.

9.4.5　小结

符号检验是作为一种非参数检验用以检验未知分布的中位数假设而被引入的. Wilcoxon-Mann-Whitney 秩检验方法是作为另一种非参数检验用以检验两个密度函数相等的假设而发展起来的. Wilcoxon-Mann-Whitney 秩检验方法被设计成当两个分布中有一个随机地大于另一个的情况下拥有较大功效函数.

习题

1. 假设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 是独立同分布的随机变量对且具有一个连续的联合分布. 令 $p=\Pr(X_i\leqslant Y_i)$，假定希望检验如下的假设：
$$H_0: p\leqslant 1/2, \\ H_1: p>1/2. \tag{9.4.7}$$
描述能检验以上假设的一个符号检验方法.
2. 重新考虑例9.4.2中的数据. 对于这两个样本应用 Kolmogorov-Smirnov 方法来检验假设(9.4.2.).
3. 重新考虑例9.4.2中的数据. 假设观察数据是取自两个方差相同的正态分布，检验假设(9.4.2)，并应用在8.4节中所介绍的 t 检验方法.

593

4. 在一个比较两种药物 A 和 B 在降低血液葡萄糖浓度作用的实验中，药物 A 被用于25个病人，药物 B 被用于15个病人. 接受药物 A 的25个病人的血液葡萄糖浓度减少的数据如表9-20所示；接受药物 B 的15个病人的血液葡萄糖浓度减少的数据如表9-21所示. 使用 Wilcoxon-Mann-Whitney 秩检验假设：这两种药物在减少血液葡萄糖浓度方面具有相同的效用.

表 9-20 在习题 4 中使用药物 A 的病人数据

0.35	0.16	0.58	1.64	1.50
1.12	1.20	0.04	0.49	1.73
1.54	0.40	0.44	0.90	1.15
0.13	1.38	0.75	0.83	0.72
0.77	0.39	0.71	0.28	0.91

表 9-21 在习题 4 中使用药物 B 的病人数据

1.78	1.82	0.68	0.89	1.26
1.25	1.95	1.48	0.86	1.07
1.01	1.81	1.59	1.63	1.31

5. 重新考虑习题 4 中的数据. 对两样本应用 Kolmogorov-Smirnov 方法检验两种药物具有相同效果的假设.
6. 重新考虑习题 4 中的数据. 假定观察数据是取自两个方差相同的正态分布. 应用在 8.4 节中所介绍到的 t 检验方法检验假设：两种药物具有相同效果.
7. 假设 $X_1,\cdots,X_m$ 是来自于概率密度函数 $f(x)$ 未知的连续分布的 m 个观察值组成的一个随机样本；$Y_1,\cdots,Y_n$ 取自另一个概率密度函数 $g(x)$ 也未知的连续分布的 n 个观察值，构成一个独立随机样本. 同时也假定对于 $-\infty<x<\infty$，都有 $f(x)=g(x-\theta)$，此处参数 $\theta(-\infty<\theta<\infty)$ 的值未知. 令 F^{-1} 为 X_i 的分位数函数，且令 G^{-1} 为 Y_i 的分位数的函数. 证明：对于所有的 $0<p<1$ 有 $F^{-1}(p)=\theta+G^{-1}(p)$.
8. 重新考虑习题 7 中的情况. 描述如何应用单边 Wilcoxon-Mann-Whitney 秩检验方法来检验如下假设：

$$H_0:\theta\leqslant 0,$$
$$H_1:\theta>0.$$

9. 重新考虑习题 7 中的条件. 描述对 θ_0 的一个特定的值，如何应用双边 Wilcoxon-Mann-Whitney 秩检验方法来检验如下的假设：

$$H_0:\theta=\theta_0,$$
$$H_1:\theta\neq\theta_0.$$

10. 重新考虑习题 9 中的情况. 描述如何使用 Wilcoxon-Mann-Whitney 秩检验方法来确定置信系数为 $1-\alpha_0$ 的 θ 的一个置信区间. 提示：当 θ_0 为何值时，在显著性水平 α_0 下，会接受原假设 $H_0:\theta=\theta_0$.
11. 令 $X_1,\cdots,X_m$ 及 $Y_1,\cdots,Y_n$ 是两个样本中的观察值，并设这些观察值中没有两个是相等的. 考虑 mn 对

$$\begin{gathered}(X_1,Y_1),\cdots,(X_1,Y_n),\\(X_2,Y_1),\cdots,(X_2,Y_n),\\\vdots\\(X_m,Y_1),\cdots,(X_m,Y_n).\end{gathered}$$

令 U 表示这些数对中 X 分量的值大于 Y 分量的值的个数. 证明：$U=S-(1/2)m(m+1)$.

其中 S 为本节所定义的分配给 $X_1,\cdots,X_m$ 的秩和.

12. 令 $X_1,\cdots,X_m$ 为具有分布函数 F 的独立同分布的随机变量，$Y_1,\cdots,Y_n$ 是具有分布函数 G 的独立同分布随机变量，且 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 互相独立，令 S 如本节所定义. 证明：(9.4.5)式给出了 S 的均值.

13. 在习题 12 的条件下，证明：(9.4.6)式给出了 S 的方差.

14. 在习题 12 和 13 的条件下，设 $F=G$. 证明 (9.4.5)式与(9.4.6)式分别和 (9.4.3)式与

594

(9.4.4)式是一致的.

15. 重新考虑习题 1 中的条件. 这次，令 $D_i=X_i-Y_i$. Wilcoxon(1945)推导了假设(9.4.7)的如下检验. 将绝对值 $|D_1|,\cdots,|D_n|$从小到大排序，且对数据分配 1 到 n 的秩. 然后令 S_W 等于那些 $D_i>0$ 的所有 $|D_i|$的秩和. 如果 $p=\Pr(X_i\leqslant Y_i)=1/2$，则 S_W 的均值和方差为

$$E(S_W)=n(n+1)/4. \tag{9.4.8}$$

$$\mathrm{Var}(S_W)=n(n+1)(2n+1)/24. \tag{9.4.9}$$

当 $S_W\geqslant c$ 时检验拒绝 H_0，这里选择的 c 使检验的显著性水平为 α_0. 称此方法为Wilcoxon 符号秩检验. 如果 n 很大，正态分布的渐近性允许我们用：

$$c=E(S_W)+\Phi^{-1}(1-\alpha_0)\mathrm{Var}(S_W)^{1/2}.$$

a. 如果 $X_i\leqslant Y_i$，令 $W_i=1$，否则令 $W_i=0$. 证明：$S_W=\sum_{i=1}^{n} iW_i$.

b. 证明在 $p=1/2$ 的假设下，$E(S_W)$如 (9.4.8)式中所述.

c. 证明：在 $p=1/2$ 的假设下，$\mathrm{Var}(S_W)$如 (9.4.9)式中所述.

16. 在一个比较可以用于制造男鞋鞋跟的两种不同的材料 A，B 的试验中，挑选了 15 位男士试穿一双用 A 材料制成一个后跟、用 B 材料制成另一个后跟的鞋子. 在试验初期，每个鞋跟有 10 毫米厚. 在鞋被穿了一个月之后，逐一测量剩余的后跟的厚度. 结果如表 9-22 所示. 检验原假设：材料 A 不如材料 B 耐用. 相对地，备择假设：材料 A 比材料 B 耐用. 使用 a. 习题 1 中的符号检验；b. 习题 15 中的 Wilcoxon 符号秩检验，c. 成对 t 检验法.

表 9-22　习题 16 中的数据

对	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
材料 A	6.6	7.0	8.3	8.2	5.2	9.3	7.9	8.5	7.8	7.5	6.1	8.9	6.1	9.4	9.1
材料 B	7.4	5.4	8.8	8.0	6.8	9.1	6.3	7.5	7.0	6.6	4.4	7.7	4.2	9.4	9.1

595

第 10 章　线性统计模型

10.1　最小二乘法

当每次由实验中得到的观察值都是成对的一组数，那么试着用其中一个数来预测另一个数通常是非常重要的．最小二乘法是一种利用观察到的成对样本从一个变量来得到另一个变量的预测方法．

10.1.1　拟合一条直线

假设给 10 个病人中每位先服用一定量的标准药物 A 来治疗，再用等量的新药物 B 治疗，然后观察每位病人对每种药的血压的变化，称这种血压的变化为病人的反应．对 $i=1,\cdots,10$，令 x_i 表示第 i 个病人服用药物 A 的反应，用适当的单位度量，用 y_i 表示第 i 个病人服用药物 B 的反应．假设表 10-1 中给出了观察到的结果．将这 10 个点 (x_i,y_i)，$i=1,\cdots,10$ 描绘在图 10-1 上．

表 10-1　两种药物的反应

i	x_i	y_i	i	x_i	y_i
1	1.9	0.7	6	4.4	3.4
2	0.8	−1.0	7	4.6	0.0
3	1.1	−0.2	8	1.6	0.8
4	0.1	−1.2	9	5.5	3.7
5	−0.1	−0.1	10	3.4	2.0

假设我们现在对描述病人对于药物 B 的反应和药物 A 的反应之间的关系感兴趣．为了得到这种关系的一个简单表达式，希望将图 10-1 上的 10 个点拟合成一条直线．尽管这 10 个点明显得并不恰好在同一条直线上，我们可能会认为该直线上的偏差是由于观察到每个病人血压的变化不仅受到两种药物的影响还受到其他因素的影响而造成的．换言之，也许相信如果可能控制其他这些因素的话，那么观察到的这 10 个点将确实在同一条直线上．还可以进一步认为，如果对大量的病人进行这两种药物的测试并观察反应结果，而不是仅仅 10 个病人，那么会发现得到的点将趋向一条直线．也许我们也希望可以通过一个病人对药物 A 的反应为 x 的基础上来预测使用药物 B 的结果 y．做这样一个预测的步骤是把图 10-1 中的点拟合成一条直线，利用该直线来预测与每个 x 值相对应的 y 的值．

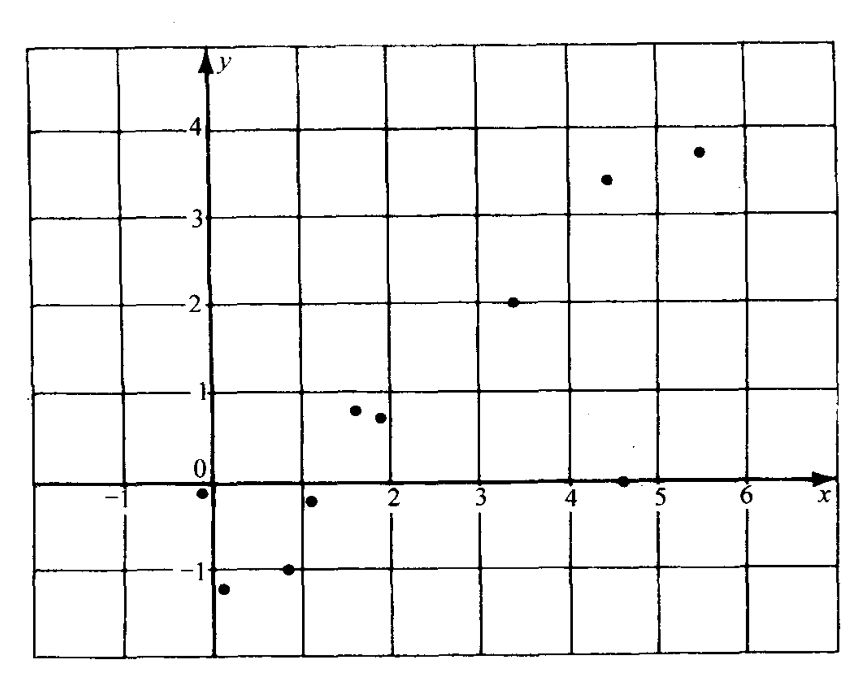


图 10-1　表 10-1 中观察数据的图

从图 10-1 中可以看出，如果我们不考虑在表 10-1 中由 $i=7$ 的病人所得到的点 (4.6,0.0)，那么其余九个点粗略地位于一条直线上. 图 10-2 中画出了对这九个点拟合得非常好的一条直线. 然而，如果希望有一条直线对这 10 个点都能适合，不清楚应该对图 10-2 中的线调整多少才能够符合这些不规则的点. 现在来描述一种能拟合该直线的方法.

600

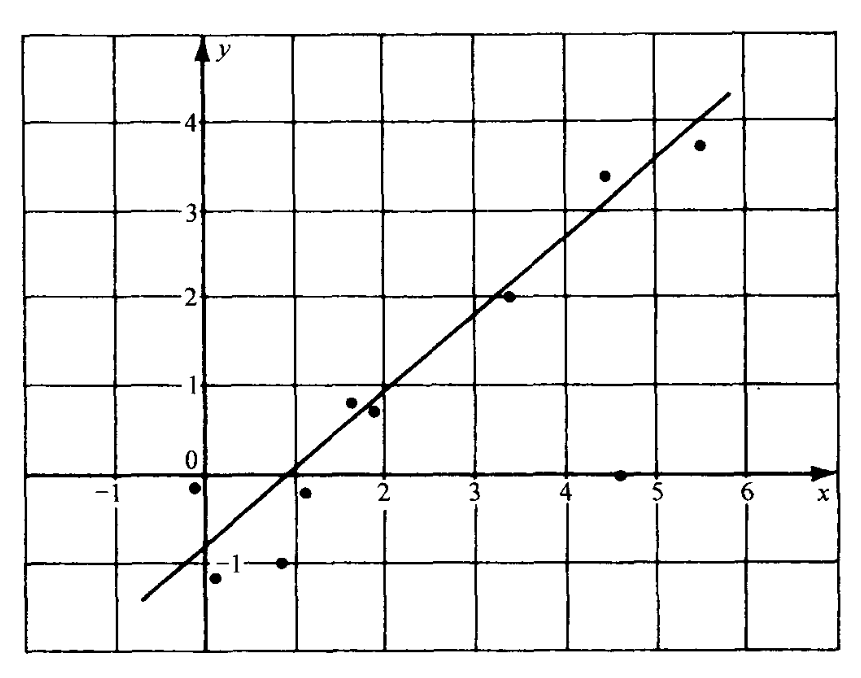


图 10-2　拟合表 10-1 中九个点的一条直线

10.1.2 最小二乘法

假设我们对于图 10-1 中所画的拟合所有点的一条直线感兴趣，为了得到把病人对新药 B 的反应 y 表示成她对普通药物 A 反应 x 的函数的一个简单数学关系式. 即主要目的是能够从她对药物 A 的反应 x 中近似地预测出对药物 B 的反应 y. 因此，我们有兴趣的是对每一个观察到的反应 x_i，构造一条直线使得直线上相应的 y 值尽可能地接近于实际上观察到的反应 y_i. 图 10-2 中所画的线到 10 个点的垂直偏差在图 10-3 中给出.

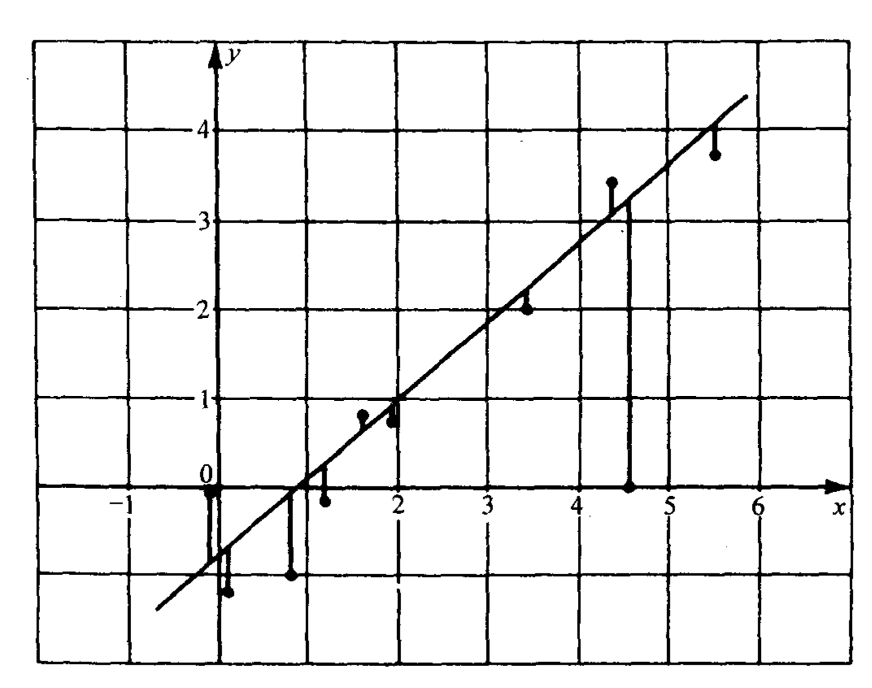


图 10-3 所画点到直线的垂直偏差

该种构造一条直线来拟合观察值的方法叫做最小二乘法. 根据该方法，这条线应该使得所有点到该线的垂直偏差的平方和达到最小值. 现在要更详细地研究这个方法.

考虑任意一条直线 $y=\beta_0+\beta_1 x$，其中常数 β_0 和 β_1 待定. 当 $x=x_i$ 时，这条线的高度是 $\beta_0+\beta_1 x_i$. 因此，点 (x_i, y_i) 和这条线之间的垂直距离是 $|y_i-(\beta_0+\beta_1 x_i)|$. 假设要用这条直线来拟合 n 个点，令 Q 表示这 n 个点的垂直距离的平方和. 那么

$$Q=\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1 x_i)]^2. \tag{10.1.1}$$

601

使用最小二乘法确定使 Q 的值达到最小而必须选择的 β_0 和 β_1 的取值.

通过考虑 β_0 和 β_1 来使 Q 的值达到最小并不难. 我们有

$$\frac{\partial Q}{\partial \beta_0}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i) \tag{10.1.2}$$

和

$$\frac{\partial Q}{\partial \beta_1}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)x_i. \tag{10.1.3}$$

通过使以上两式偏导数等于0，得到以下一对方程：

$$\beta_0 n+\beta_1 \sum_{i=1}^{n} x_i=\sum_{i=1}^{n} y_i,$$
$$\beta_0 \sum_{i=1}^{n} x_i+\beta_1 \sum_{i=1}^{n} x_i^2=\sum_{i=1}^{n} x_i y_i. \tag{10.1.4}$$

称方程(10.1.4)为β_0和β_1的正规方程. 通过考虑Q的二阶偏导数，可以证明满足正规方程的β_0和β_1的值是使方程(10.1.1)中平方和Q最小的值. 如果用$\hat{\beta}_0$和$\hat{\beta}_1$来代表这些值，那么通过最小二乘法得到的直线方程将会是$y=\hat{\beta}_0+\hat{\beta}_1 x$. 该条直线称作最小二乘直线.

602

通常，令$\bar{x}_n=(1/n)\sum_{i=1}^{n} x_i$和$\bar{y}_n=(1/n)\sum_{i=1}^{n} y_i$. 对$\beta_0$和$\beta_1$通过求解(10.1.4)中的正规方程，得到以下的结果：

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n} x_i y_i-n\bar{x}_n\bar{y}_n}{\sum_{i=1}^{n} x_i^2-n\bar{x}_n^2},$$
$$\hat{\beta}_0=\bar{y}_n-\hat{\beta}_1\bar{x}_n. \tag{10.1.5}$$

对表10-1中给出的值，$n=10$，由(10.1.5)式可以得到$\hat{\beta}_0=-0.786$, $\hat{\beta}_1=0.685$. 因此最小二乘直线的方程是$y=-0.786+0.685x$. 这条线见图10-4中所画.

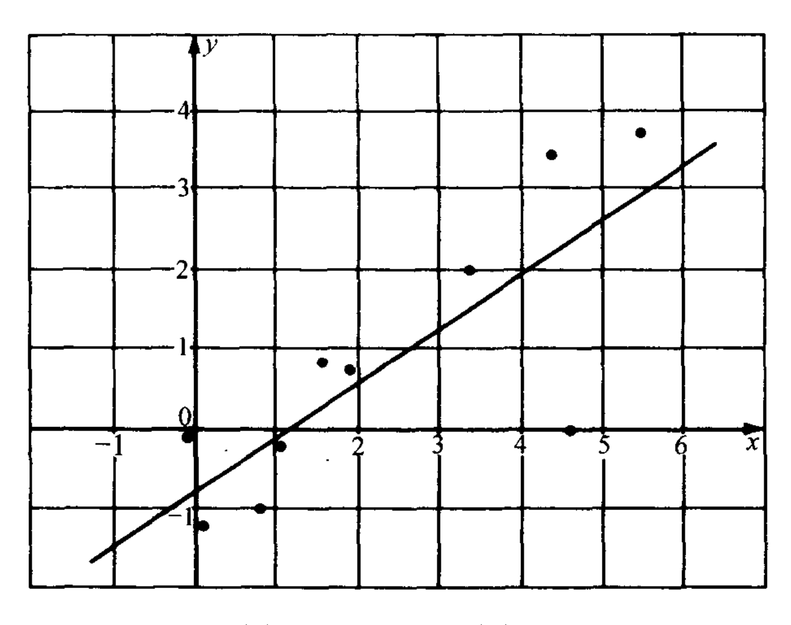


图10-4　最小二乘直线

事实上用所有计算机统计软件都可以计算出最小二乘回归直线. 甚至一些便携式计算器也可以做这样的计算.

10.1.3　用最小二乘法来拟合一个多项式

现在假设不是简单地构造一条直线来拟合所画的n个点，而是要构造一个$k(k\geqslant 2)$

阶多项式. 这样一个多项式将有以下形式:

$$y=\beta_0+\beta_1 x+\beta_2 x^2+\cdots+\beta_k x^k. \tag{10.1.6}$$

最小二乘法会使从曲线到点的垂直偏差的平方和 Q 达到最小, 从而来确定常数 $\beta_0,\cdots,\beta_k$, 即, 应该选择使得以下 Q 的表达式最小化的这些常数:

603

$$Q=\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1 x_i+\cdots+\beta_k x_i^k)]^2. \tag{10.1.7}$$

如果计算 $k+1$ 个偏导数 $\frac{\partial Q}{\partial\beta_0},\cdots,\frac{\partial Q}{\partial\beta_k}$, 并且令每个偏导数等于0, 就得到以下含有 $k+1$ 个未知数 $\beta_0,\cdots,\beta_k$ 的 $k+1$ 个线性方程:

$$\begin{aligned}
\beta_0 n+\beta_1\sum_{i=1}^{n}x_i+\cdots+\beta_k\sum_{i=1}^{n}x_i^k&=\sum_{i=1}^{n}y_i,\\
\beta_0\sum_{i=1}^{n}x_i+\beta_1\sum_{i=1}^{n}x_i^2+\cdots+\beta_k\sum_{i=1}^{n}x_i^{k+1}&=\sum_{i=1}^{n}x_i y_i,\\
&\vdots\\
\beta_0\sum_{i=1}^{n}x_i^k+\beta_1\sum_{i=1}^{n}x_i^{k+1}+\cdots+\beta_k\sum_{i=1}^{n}x_i^{2k}&=\sum_{i=1}^{n}x_i^k y_i.
\end{aligned} \tag{10.1.8}$$

与前相同, 称该方程组为正规方程组. 存在唯一一组 $\beta_0,\cdots,\beta_k$ 满足正规方程组, 当且仅当由该方程组中 $\beta_0,\cdots,\beta_k$ 的系数所构成 $(k+1)\times(k+1)$ 阶矩阵的行列式不为零. 若 n 个观测值 $x_1,\cdots,x_n$ 中至少有 $k+1$ 个不同的值, 那么该矩阵的行列式不为零且该正规标准方程组有唯一的一组解. 假定该条件是满足的. 通过进一步的计算可以证明满足正规方程组的 $\beta_0,\cdots,\beta_k$ 的唯一的一组解将会使方程(10.1.7)中的 Q 最小化. 将解记为 $\hat\beta_0,\cdots,\hat\beta_k$, 则最小二乘多项式为 $y=\hat\beta_0+\hat\beta_1 x+\cdots+\hat\beta_k x^k$.

例 10.1.1 拟合一个抛物线 假定希望用数据表 10-1 中的 10 个点来拟合形式如 $y=\beta_0+\beta_1 x+\beta_2 x^2$ 的多项式看上去是一条抛物线. 在本例中, 可以发现(10.1.8)中的正规方程如下:

$$\begin{aligned}
10\beta_0+23.3\beta_1+90.37\beta_2&=8.1,\\
23.3\beta_0+90.37\beta_1+401.0\beta_2&=43.59,\\
90.37\beta_0+401.0\beta_1+1892.7\beta_2&=204.55.
\end{aligned} \tag{10.1.9}$$

而满足这三个方程的 β_0、β_1 和 β_2 的唯一的一组解是 $\hat\beta_0=-0.744$, $\hat\beta_2=0.616$ 和 $\hat\beta_2=0.013$. 因此, 该最小二乘抛物曲线方程为

$$y=-0.744+0.616x+0.013x^2. \tag{10.1.10}$$

图 10-5 显示了该曲线及其最小二乘直线. 由于方程(10.1.10)中 x^2 的系数很小, 所以在图 10-5 所包括的取值范围内该最小二乘抛物曲线与最小二乘直线之间都是很接近的. ◀

604

图 10-5　最小二乘抛物线

例 10.1.2　汽油英里数　Heavenrich 和 Hellman(1999)公布了测量 173 辆不同的车辆而得到的几个变量. 在这些变量中有两个变量是汽油英里数(英里/每加仑)和引擎马力. 这两个变量的散点图以及通过最小二乘法所拟合的抛物曲线如图 10-6 所示. 即使没有图 10-6 中所描绘的曲线，也明显可以看到用直线不能充分说明这两个变量之间的关系. 而某种曲线关系必定拟合得好. 图中，该最小二乘抛物曲线在变量引擎马力达到最大值时出现向上弯曲，而这多少有点出人意料. 事实上，对于这样一个例子，应该使用一些先验信息来对拟合的曲线方程加以某种限制. 另外，可以使用每加仑汽油英里数的曲线函数来代替汽油英里数，并用该曲线函数当作 y 变量. ◀

605

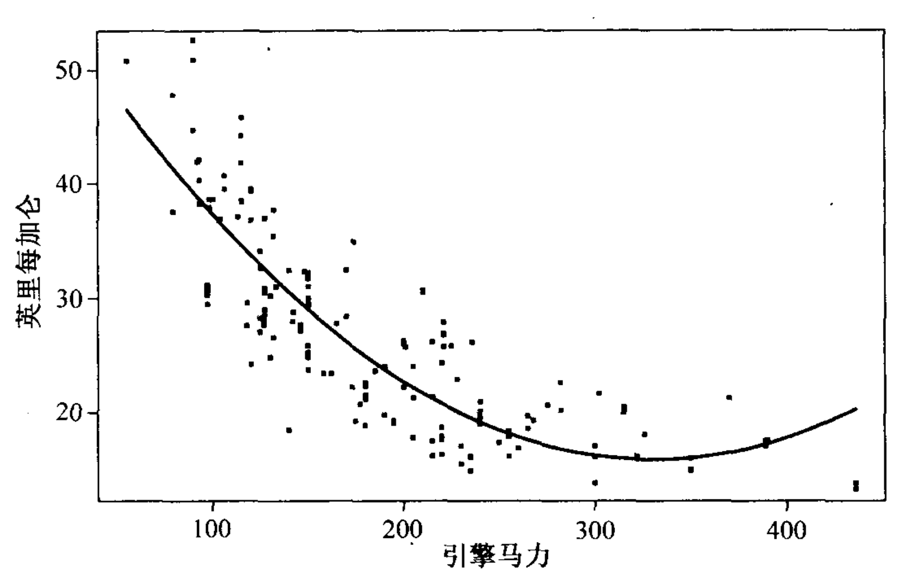


图 10-6　例 10.1.2 中 173 辆小轿车每加仑英里与引擎马力的散点图. 最小二乘抛物曲线也在图中给出

10.1.4 拟合多个变量的一个线性函数

现在将考虑在本节开始所讨论的例子的推广，在该例中我们将某病人对一种新药 B 的反应表示成一个对药物 A 的反应有关的线性函数感兴趣. 假设希望将这个病人对 B 药物的反应表示为不仅包含对药物 A 的反应，还包括其他一些相关变量的一个线性函数. 例如，希望将病人对药物 B 的反应 y 表示成包括她对药物 A 的反应 x_1、在接受任何药物前的心律 x_2 和血压 x_3，以及其他的一些相关变量 $x_4,\cdots,x_k$ 的一个线性函数.

假设对于每一个病人 $i(i=1,\cdots,n)$，测量她对药物 B 的反应 y_i，对药物 A 的反应 x_{i1}，以及其他一些变量的值 $x_{i2},\cdots,x_{ik}$. 为了拟合 n 个病人的这些观察值，考虑如下形式的一个线性函数

$$y=\beta_0+\beta_1x_1+\cdots+\beta_kx_k. \tag{10.1.11}$$

在该种情况下，可以通过最小二乘法来计算得到$\beta_0,\cdots,\beta_k$的值. 对每组给定的观察值 $x_{i1},\cdots,x_{ik}$，再次考虑被观察到的反应 y_i 与式(10.1.11)中给出的线性函数 $\beta_0+\beta_1x_1+\cdots+\beta_kx_{ik}$值之间的差距. 与前相同，需要将这些差的平方和 Q 最小化. 其中

$$Q=\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1x_{i1}+\cdots+\beta_kx_{ik})]^2. \tag{10.1.12}$$

通过计算每个偏导数 $\partial Q/\partial\beta_j,j=0,1,\cdots,k+1$，并令它们等于0，可得到 $k+1$ 个正规方程. 这些方程有如下形式：

$$\begin{aligned}
\beta_0n+\beta_1\sum_{i=1}^{n}x_{i1}+\cdots+\beta_k\sum_{i=1}^{n}x_{ik} &= \sum_{i=1}^{n}y_i,\\
\beta_0\sum_{i=1}^{n}x_{i1}+\beta_1\sum_{i=1}^{n}x_{i1}^2+\cdots+\beta_k\sum_{i=1}^{n}x_{i1}x_{ik} &= \sum_{i=1}^{n}x_{i1}y_i,\\
&\vdots\\
\beta_0\sum_{i=1}^{n}x_{ik}+\beta_1\sum_{i=1}^{n}x_{ik}x_{i1}+\cdots+\beta_k\sum_{i=1}^{n}x_{ik}^2 &= \sum_{i=1}^{n}x_{ik}y_i.
\end{aligned} \tag{10.1.13}$$

606

如果由这些方程系数$\beta_0,\cdots,\beta_k$ 所组成的 $(k+1)\times(k+1)$阶矩阵的行列式不为0，则这些方程有唯一解 $\hat{\beta}_0,\cdots,\hat{\beta}_k$. 从而最小二乘的线性函数是 $y=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k$.

例 10.1.3 拟合两个变量的一个线性函数 假设将表 10-1 扩展为包含表 10-2 中的第三列. 其中，对于每个病人 $i(i=1,\cdots,10)$，x_{i1}表示她对标准药物 A 的反应，x_{i2} 表示她的心律，y_i 表示她对新药 B 的反应. 同样假定希望拟合一个适合这些值的具有 $y=\beta_0+\beta_1x_1+\beta_2x_2$ 形式的线性函数.

在本例中，可以发现正规方程(10.1.13)为：

$$\begin{aligned}
10\beta_0+23.3\beta_1+650\beta_2&=8.1,\\
23.3\beta_0+90.37\beta_1+1563.6\beta_2&=43.59,\\
650\beta_0+1563.6\beta_1+42.334\beta_2&=563.1.
\end{aligned} \tag{10.1.14}$$

满足以上三个方程的 β_0、β_1、β_2 的唯一值为 $\hat{\beta}_0 = -11.4527$，$\hat{\beta}_1 = 0.4503$，$\hat{\beta}_2 = 0.1725$，因此最小二乘线性函数为

$$y = -11.4527 + 0.4503x_1 + 0.1725x_2. \qquad (10.1.15)$$ ◀

表 10-2　对两种药物的反应和心律

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	1.9	66	0.7	6	4.4	70	3.4
2	0.8	62	-1.0	7	4.6	68	0.0
3	1.1	64	-0.2	8	1.6	62	0.8
4	0.1	61	-1.2	9	5.5	68	3.7
5	-0.1	63	-0.1	10	3.4	66	2.0

值得注意的是拟合一个只含有一个变量的 k 次多项式，如公式(10.1.6)所示，可以认为是拟合一个如(10.1.1)式中含有多个变量的线性函数问题的一个特例. 为了使等式(10.1.11)式能适用于拟合具有(10.1.6)式所给出的多项式形式的问题，只须简单地定义 k 个变量 $x_1, \cdots, x_k$ 为 $x_1 = x, x_2 = x^2, \cdots, x_k = x^k$.

607

一个含有不止一个变量的多项式可以表示成(10.1.11)式的形式. 例如，假设四个变量 r，s，t 和 y 的值是对几个不同的病人观察到的，我们希望拟合这些观察值的函数具有如下形式：

$$y = \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 rs + \beta_4 s^2 + \beta_5 t^3 + \beta_6 rst. \qquad (10.1.16)$$

通过定义 6 个变量 $x_1, \cdots, x_6$ 如下：$x_1 = r, x_2 = r^2, x_3 = rs, x_4 = s^2, x_5 = t^3$ 以及 $x_6 = rst$，可以将(10.1.16)式中的函数看作是在 $k=6$ 时具有(10.1.11)式中所给形式的一个线性函数.

10.1.5　小结

用最小二乘法基于一个或多个变量($x_1, \cdots, x_k$)来计算一个变量(y)的具有 $\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ 形式的估计值. 应选择系数 $\beta_0, \cdots, \beta_k$ 使得 y 的观察值和 $\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ 的观察值之间差的平方和尽可能的小. 已经给出了这些系数的代数公式，但大多数计算机统计软件能更容易地算出这些系数.

习题

1. 证明：$\sum_{i=1}^{n}(c_1 x_i + c_2)^2 = c_1^2 \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + n(c_1 \bar{x}_n + c_2)^2$.

2. 证明：(10.1.5)式中的 $\hat{\beta}_1$ 的值能重写成下列三种形式中的任何一种：

a. $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$，　b. $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$，　c. $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}x_i(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$.

3. 证明：最小二乘直线 $y=\hat{\beta}_0+\hat{\beta}_1 x$ 通过点$(\bar{x}_n,\bar{y}_n)$.

4. 对于 $i=1,\cdots,n$，令 $\hat{y}_i=\beta_0+\beta_1 x_i$. 证明：(10.1.5)式中给出的$\hat{\beta}_0$ 和$\hat{\beta}_1$ 是使得 $\sum_{i=1}^{n}(y_i-\hat{y}_i)=0$ 和 $\sum_{i=1}^{n}x_i(y_i-\hat{y}_i)=0$ 成立的唯一值.

5. 拟合表 10-1 中所给观察值的一条直线，使得点到线的水平偏差的平方和最小. 在同一张图上画出这条线和图 10-4 中所给出的最小二乘直线.

6. 已知最小二乘直线和最小二乘抛物线都拟合同一个点集. 解释为什么来自抛物线的点的偏差的平方和不可能大于来自直线的点的偏差的平方和.

7. 已知某种合金的八个样本是在不同温度下生产出来的，而且观察了每个样本的韧度. 观察值列于表 10-3 中，其中 x_i 表示样本 i 生产时的温度(标准单位)，而 y_i 表示样本的韧度(标准单位).

608

   a. 用最小二乘法将这些值拟合成一条形如 $y=\beta_0+\beta_1 x$ 的直线；
   b. 用最小二乘法将这些值拟合成一条形如 $y=\beta_0+\beta_1 x+\beta_2 x^2$ 的抛物线.
   c. 在同一张图中描出这八个数据点，以及在 (a) 中得到的直线和 (b) 中得到的抛物线.

表 10-3 习题 7 的数据

i	x_i	y_i	i	x_i	y_i
1	0.5	40	5	2.5	44
2	1.0	41	6	3.0	42
3	1.5	43	7	3.5	43
4	2.0	42	8	4.0	42

8. 设 $(x_i,y_i),i=1,\cdots,k+1$，表示在 xy－平面内给定使得任意两点之间没有相同的横坐标的 $k+1$ 个点. 证明：存在唯一的具有形式 $y=\beta_0+\beta_1 x+\cdots+\beta_k x^k$ 的多项式经过这 $k+1$ 个点.

9. 将某种塑料的弹力 y 表示成塑料被浇固温度 x_1 和被浇固时间 x_2 的一个线性函数. 已知在 x_1 和 x_2 的不同数值下做出了 10 段塑料，被观察到的值以适当的单位在表 10-4 中给出. 用最小二乘法拟合一个形如 $y=\beta_0+\beta_1 x+\beta_2 x^2$ 的函数.

表 10-4 习题 9 的数据

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	100	1	113	6	120	2	144
2	100	2	118	7	120	3	138
3	110	1	127	8	130	1	146
4	110	2	132	9	130	2	156
5	120	1	136	10	130	3	149

10. 再次考虑表 10-4 中的观察值，用最小二乘法拟合一个形如 $y=\beta_1 x_1+\beta_2 x_2+\beta_3 x_2^2$ 的函数.

11. 再次考虑表 10-4 中的观察值，思考习题 9 和 10 中的拟合这些变量的两个函数. 问哪个函数更好地拟合了观察值.

10.2 回归分析

在10.1节中介绍了最小二乘法的方法. 这种用来预测变量y的方法是基于其他变量$x_1,\cdots,x_k$基础上计算线性函数的系数. 在本节中，假设y的值是一族随机变量的观察值. 在这种情况下，有一个统计模型使得应用最小二乘法的结果产生了该模型参数的极大似然估计.

10.2.1 回归系数

在本节中，先介绍一些使用最小二乘法来得到各种参数估计值的统计问题. 特别地，应该研究这样一些问题，我们对于在给定其他变量$X_1,\cdots,X_k$值的条件下某个随机变量Y的条件分布感兴趣. 这些变量$X_1,\cdots,X_k$可能是随机变量，在试验中它们的值和Y的值都是待观察的，或者它们可能是控制变量，其值由试验者来选择. 一般地，其中一些变量可能是随机变量，而另一些可能是控制变量. 但无论如何，我

609

们要研究的是在给定$X_1,\cdots,X_k$的条件下的Y的条件分布.

在给定$X_1,\cdots,X_k$的值的条件下Y的条件期望被称作Y关于$X_1,\cdots,X_k$的回归函数，或者简单地称作Y关于$X_1,\cdots,X_k$的回归. Y关于$X_1,\cdots,X_k$的回归是关于$X_1,\cdots,X_k$的值$x_1,\cdots,x_k$的一个函数，该函数是$\mathrm{E}(Y|x_1,\cdots,x_k)$.

假设回归函数$\mathrm{E}(Y|x_1,\cdots,x_k)$是具有如下形式的一个线性函数：

$$\mathrm{E}(Y|x_1,\cdots,x_k)=\beta_0+\beta_1x_1+\cdots+\beta_kx_k. \tag{10.2.1}$$

在公式(10.2.1)中的系数$\beta_0,\cdots,\beta_k$称为回归系数. 假定这些回归系数是未知的. 因此，它们是要估计的参数. 还假定可以得到观察值的n组向量值. 对于从$i=1,\cdots,n$，假设第i组向量$(x_{i_1},\cdots,x_{i_k},y_i)$是由$X_1,\cdots,X_k$的控制值或观察值及与之相对应的$Y$的观测值所组成.

根据10.1节中所描述的由这些观察值计算出来的回归系数$\beta_0,\cdots,\beta_k$的一组估计值是由最小二乘法得到的一组值$\hat{\beta}_0,\cdots,\hat{\beta}_k$. 称这些估计值为$\beta_0,\cdots,\beta_k$的最小二乘估计值. 现在我们对于给定$X_1,\cdots,X_k$值条件下$Y$的条件分布做一些进一步的假设，目的是为了能够确定这些最小二乘估计值更为详细的性质.

10.2.2 简单线性回归

首先来考虑一个仅研究Y对于单个变量X的回归情况. 假设对于每个$X=x$的值，随机变量Y可以表示成$y=\beta_0+\beta_1x+\varepsilon$的形式，这里的$\varepsilon$是一个服从均值为0、方差为$\sigma^2$的正态分布的随机变量. 由假设可知对于给定的$X=x$，$Y$的条件分布为均值是$\beta_0+\beta_1x$、方差是$\sigma^2$的正态分布.

该类问题称为简单线性回归. 此处的简单是指我们仅考虑关于一个变量X时Y的回归情况，而不是基于多个变量，线性是指回归方程$\mathrm{E}(Y|x)=\beta_0+\beta_1x$是参数为

β_0 和 β_1 的一个线性方程. 例如，一个问题中的 $E(Y|x)$ 是一个多项式，就像(10.1.6)中右边的式子，也是一个线性回归问题，但不是简单的.

现在假设我们得到了 n 对观测值 $(x_1,Y_1),\cdots,(x_n,Y_n)$，并且假设：$x_1,\cdots,x_n$ 的值都事先知道，或者在计算 $(Y_1,\cdots Y_n)$ 的联合分布之前就先以他们的值为条件. 随机变量 $Y_1,\cdots Y_n$ 是相互独立的，且对于 $i=1,\cdots,n$，Y_i 的分布是一个均值为 $\beta_0+\beta_1 x$ 和方差为 σ^2 的正态分布. 因此，在给定向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$ 的值和 β_0、β_1 和 σ^2 的条件下，$Y_1,\cdots,Y_n$ 的联合概率密度函数为：

$$f_n(\boldsymbol{y}|\boldsymbol{x},\beta_0,\beta_1,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2\right]. \quad (10.2.2)$$

610

对于每个观察到的向量 $\boldsymbol{y}=(y_1,\cdots,y_n)$，该函数将会是参数 β_0、β_1 和 σ^2 的似然函数. 现在可以用该似然函数来确定 β_0、β_1 和 σ^2 的极大似然估计. 从方程(10.2.2)中可以看出不管 σ^2 的值是什么，似然函数达到最大值时的 β_0、β_1 的值是使下式的平方和取最小的值：

$$\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2.$$

但是该平方和(10.1.1)中给出的平方和 Q 是完全一样的，通过最小二乘法可以使其取到最小值. 因此，回归系数 β_0、β_1 的极大似然估计值也就是用最小二乘法估算的 β_0 和 β_1. 这些估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的确切形式已在等式(10.1.5)中给出.

最后，首先用 β_0、β_1 的极大似然估计值 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 来代替方程(10.2.2)中的 β_0 和 β_1，然后把关于 σ^2 的结果表达式最大化便可求得 σ^2 的极大似然估计. 该结果（参见这节后面的习题1）是：

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\beta}_0+\hat{\beta}_1x_i)^2. \quad (10.2.3)$$

除了假设 Y 关于 X 的回归是形式如 $E(Y|x)=\beta_0+\beta_1 x$ 的线性方程之外，还需进一步对于 $Y_1,\cdots,Y_n$ 的关于所有 $x_1,\cdots,x_n$ 值的联合分布做三个假设. 这些假设可以总结如下：

i. *正态性*. 假设每个变量 Y_i 都服从正态分布.

ii. *独立性*. 假定变量 $Y_1,\cdots,Y_n$ 是互相独立的.

iii. *同方差性*. 假定变量 $Y_1,\cdots,Y_n$ 具有相同的方差 σ^2. 称这个假设为*同方差性*假设. 一般地，当随机变量具有相同方差时称为*同方差性*，随机变量有不同方差时称为*异方差性*.

在这些假设下我们可以证明 β_0 和 β_1 的极大似然估计是由方程(10.1.5)所确定的，而 σ^2 的极大似然估计则由方程(10.2.3)给定.

10.2.3 最小二乘估计的分布

现在要讨论对于给定 $x_1,\cdots,x_n$ 的值，把 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 这两个估计值看作随机变量

$Y_1,\cdots,Y_n$的函数时的联合分布. 应该记住只要先前描述的三个假设成立，则不管是在 $x_1,\cdots,x_n$ 的值事实上是实验者所选择的问题中，还是在 $x_1,\cdots,x_n$ 的值和 $Y_1,\cdots,Y_n$ 的

611 值是一起被观察到的问题中，这里的分析都是正确的.

为方便起见，对本节和下一节内容，引入符号

$$s_x = \left(\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right)^{1/2}. \tag{10.2.4}$$

为了确定 $\hat{\beta}_1$ 的分布，将 $\hat{\beta}_1$ 写成下面形式会更方便些(参看10.1节最后的习题2)：

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)Y_i}{s_x^2}. \tag{10.2.5}$$

由(10.2.5)式可以看出 $\hat{\beta}_1$ 是关于 $Y_1,\cdots,Y_n$ 的一个线性函数. 因为随机变量 $Y_1,\cdots,Y_n$ 是相互独立且每个都服从正态分布，因而 $\hat{\beta}_1$ 也服从正态分布. 进一步地，该分布的均值为

$$\mathrm{E}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)\mathrm{E}(Y_i)}{s_x^2}.$$

因为 $\mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i$，当 $i=1,\cdots,n$，可以发现(参看本节最后部分的习题2)

$$\mathrm{E}(\hat{\beta}_1) = \beta_1. \tag{10.2.6}$$

因此，$\hat{\beta}_1$ 是 β_1 的无偏估计值.

进一步地，由于随机变量 $Y_1,\cdots Y_n$ 是相互独立的并且每个都有方差 σ^2，由等式(10.2.5)式可得

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\mathrm{Var}(Y_i)}{s_x^4} = \frac{\sigma^2}{s_x^2}. \tag{10.2.7}$$

现在来考虑 $\hat{\beta}_0$ 的分布. 由(10.1.5)式可得 $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1\bar{x}_n$，因为 $\bar{Y}_n$ 和 $\hat{\beta}_1$ 都是 $Y_1,\cdots,Y_n$ 的线性函数，可知 $\hat{\beta}_0$ 也是 $Y_1,\cdots,Y_n$ 一个线性函数. 因此 $\hat{\beta}_0$ 服从正态分布. $\hat{\beta}_0$ 的均值可以由关系式 $\mathrm{E}(\hat{\beta}_0) = \mathrm{E}(\bar{Y}_n) - \bar{x}_n\mathrm{E}(\hat{\beta}_1)$来确定. 可以证明(参看习题3)

$$\mathrm{E}(\hat{\beta}_0) = \beta_0. \tag{10.2.8}$$

因此，$\hat{\beta}_0$ 是 $\hat{\beta}_0$ 的无偏估计值.

此外，还可以证明(参看习题4)

612

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_x^2}\right). \tag{10.2.9}$$

接下来，考虑 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的协方差. 可以证明这个协方差的值可以表示为如下形式(参见习题5)

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}_n^2\sigma^2}{s_x^2}. \tag{10.2.10}$$

通过进一步的学习，我们会知道β_0和β_1的联合分布是一个服从由(10.2.6)式至(10.2.10)式中所给出的均值为$E(\hat{\beta}_0)$和$E(\hat{\beta}_1)$、方差为$\mathrm{Var}(\hat{\beta}_0)$和$\mathrm{Var}(\hat{\beta}_1)$、以及协方差为$\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)$的二元正态分布.

例 10.2.1 压强和水的沸点 福布斯(Forbes)(1857)报道了一个为得到高度估计值的方法而进行的试验所得出的结论. 高度和大气压强之间有个公式. 但是在福布斯时代要把气压计带到很高的地方是很困难的. 然而，对旅游者来说带一个温度计并测量水的沸点有可能是容易的. 表 10-5 中包含了从 17 个试验中所测得的大气压强和水的沸点. 我们可以用简单线性回归来估计沸点和压强之间的关系. 令Y_i表示福布斯观察结果中的大气压强，令x_i是相应的沸点，其中$i=1,\cdots,17$. 假设Y_i有均值 613 $\beta_0+\beta_i x_i$和方差σ^2，且相互独立. 当$n=17$时，平均温度$\bar{x}_n=202.95$且$s_x^2=530.78$. 从这些值中，我们可以用本节中得到的公式计算出最小二乘估计量的方差和协方差. 例如：

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{530.78}=0.00188\sigma^2,$$

$$\mathrm{Var}(\hat{\beta}_0)=\sigma^2\left(\frac{1}{17}+\frac{202.95^2}{530.78}\right)=77.66\sigma^2,$$

$$\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_0)=\frac{202.95^2}{530.78}=0.382\sigma^2.$$

容易看出我们得出β_1的估计值要比β_0的估计值精确得多. ◀

表 10-5 由福布斯试验得到的华氏温度下水的沸点和由水银英寸所表示的大气压强. 这些数据来自 Weisberg (1985, p.3)

沸点	压强	沸点	压强	沸点	压强
194.5	20.79	200.9	23.89	209.5	28.49
194.3	20.79	201.1	23.99	208.6	27.76
197.9	22.40	201.4	24.02	210.7	29.04
198.4	22.67	201.3	24.01	211.9	29.88
199.4	23.15	203.6	25.14	212.2	30.06
199.9	23.35	204.6	26.57		

例 10.2.2 线性组合的方差 我们经常需要计算最小二乘估计的线性组合的方差. 在本节后面要讨论的一个预测问题中有这样的例子. 假如想要计算$T=c_0\hat{\beta}_0+c_1\hat{\beta}_1+c_*$的方差. 把(10.2.9)、(10.2.7)和(10.2.10)式中给出的$\mathrm{Var}(\hat{\beta}_0)$，$\mathrm{Var}(\hat{\beta}_1)$和$\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_0)$代入以下关系式就可以得到$T$的方差：

$$\mathrm{Var}(T)=c_0^2\mathrm{Var}(\hat{\beta}_0)+c_1^2\mathrm{Var}(\hat{\beta}_1)+2c_0c_1\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_0).$$

把这些方差代入公式后，这个结果可以重写成如下形式：

$$\mathrm{Var}(T)=\sigma^2\left(\frac{c_0^2}{n}+\frac{(c_0\bar{x}_n-c_1)^2}{s_x^2}\right). \qquad (10.2.11)$$ ◀

10.2.4　预测

假设在一个简单线性回归问题中得到了 n 对观察值 $(x_1, Y_1), \cdots, (x_n, Y_n)$，在这 n 对数据的基础上，当赋予控制变量某个确定的值 x，就能得到 Y 时，则有必要预测独立观察量 Y 的值. 由于观察值 Y 服从均值为 $\beta_0 + \beta_1 x$ 和方差为 σ^2 的一个正态分布，很自然地就会用 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 作为 Y 的预测值. 现在要确定该预测的均方误差 $E[(\hat{Y} - Y)^2]$，其中 $\hat{Y}$ 和 Y 都是随机变量.

在该问题中，$E(\hat{Y}) = E(Y) = \beta_0 + \beta_1 x$. 因此，如果令 $\mu = \beta_0 + \beta_1 x$，那么

$$E[(\hat{Y} - Y)^2] = E\{[(\hat{Y} - \mu) - (\hat{Y} - \mu)]^2\} \tag{10.2.12}$$

614

$$= \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y) - 2\mathrm{Cov}(\hat{Y}, Y).$$

然而，随机变量 $\hat{Y}$ 和 Y 是独立的，因为 $\hat{Y}$ 是前 n 对观察值的一个函数，而 Y 是一个独立的观察值，所以 $\mathrm{Cov}(\hat{Y}, Y) = 0$，因此就有：

$$E[(\hat{Y} - Y)^2] = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y). \tag{10.2.13}$$

最后，因为 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$，当 $c_0 = 1, c_1 = x$ 时，$\mathrm{Var}(\hat{Y})$ 的值由(10.2.11)式所给定. 因为 $\mathrm{Var}(\hat{Y}) = \sigma^2$，我们有：

$$E[(\hat{Y} - Y)^2] = \sigma^2\left[1 + \frac{1}{n} + \frac{(x - \bar{x}_n)^2}{s_x^2}\right]. \tag{10.2.14}$$

例 10.2.3　从水的沸点预测压强　在例 10.2.1 中，福布斯正试图用水的沸点来估计大气压强. 利用表 10-5 里的数据，我们可以计算出简单线性回归参数的最小二乘估计：$\hat{\beta}_0 = -81.049$，$\hat{\beta}_1 = 0.5228$，$\hat{\sigma}^2 = 0.0478$. 图 10-7 表明所描绘的数据和最小二乘回归直线. 现在，假设一个旅行者测得水的沸点是 201.5 度. 压强 Y 预测值的均方误差可以从(10.2.14)式中获得.

$$E[(\hat{Y} - Y)^2] = \sigma^2\left[1 + \frac{1}{17} + \frac{(201.5 - 202.95)^2}{530.78}\right] = 1.0628\sigma^2,$$

其中预测的观察值是 $\hat{Y} = -81.06 + 0.5229 \times 201.5 = 24.30$. 在图 10-7 中标明了 $\hat{Y}$ 的计算结果. 可以如下地解释均方误差 $1.0628\sigma^2$：如果我们知道 β_0 和 β_1 的值并试图预测 Y，均方误差会是 $\mathrm{Var}(Y) = \sigma^2$. 要估计 β_0 和 β_1，使得我们付出在均方误差中增加了额外的 $0.0628\sigma^2$ 的代价. ◀

615

注： 当 x 远离观察数据时，预测值的均方误差会增加. 当 x 偏离 $\bar{x}_n$ 的距离增大时，(10.2.14)式中的均方误差也会增加，并且当 $x = \bar{x}_n$ 时，均方误差达到最小. 这表明当 x 不在观察值 $x_1, \cdots, x_n$ 的中心附近时，要预测 Y 是困难的. 事实上，如果 x 比最大的观察值 x_i 更大或者比最小的更小时，要更精确地预测 Y 都是相当困难的. 这种在观察数据范围之外的预测称作外推.

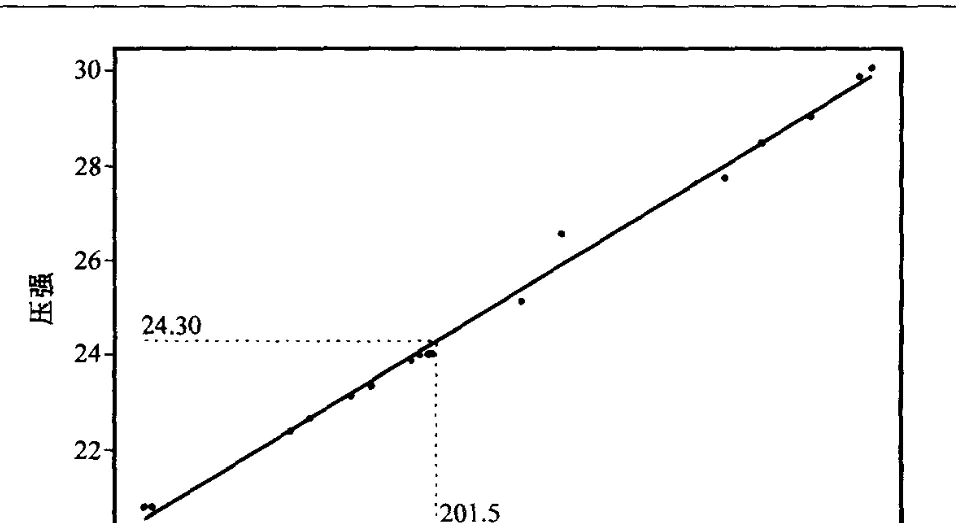


图 10-7 例 10.2.3 中压强对沸点的图及其回归直线.
虚线说明当沸点为 201.5 时压强的预测值

10.2.5 扩展研究：试验设计

考虑一个简单的线性回归问题，其中变量 X 是一个受控制变量，它的值 $x_1,\cdots,x_n$ 可以由实验者来选择. 我们要讨论为了得到回归系数 β_0 和 β_1 的估计量而选择这些值的方法.

首先假设要选择 $x_1,\cdots,x_n$ 的值使得最小二乘估计量 $\hat{\beta}_0$ 的均方误差最小化. 因为 $\hat{\beta}_0$ 是 β_0 的无偏估计量，所以 $\hat{\beta}_0$ 的均方差等于如式(10.2.9)所给出的 $\hat{\beta}_0$ 的方差. 由(10.2.9)式可得对于所有值 $x_1,\cdots,x_n$ 都有 $\mathrm{Var}(\hat{\beta}_0)\geqslant\sigma^2/n$，而且当且仅当 $\bar{x}_n=0$ 时上述关系式中的等号才成立，所以当 $\bar{x}_n=0$ 时，$\mathrm{Var}(\hat{\beta}_0)$ 取到其最小值 σ^2/n.

再假设要选取值 $x_1,\cdots,x_n$ 使得估计量 $\hat{\beta}_1$ 的均方差最小化. 再次取 $\hat{\beta}_1$ 的均方差等于 $\mathrm{Var}(\hat{\beta}_1)$. 由(10.2.7)式可以看到通过选择使得 s_x^2 达到最大的值 $x_1,\cdots,x_n$ 可以使 $\mathrm{Var}(\hat{\beta}_1)$ 达到最小. 如果 $x_1,\cdots,x_n$ 取自实数轴上的某个有界区间 (a,b)，且如果 n 是一个偶数，那么恰好 $n/2$ 的值取 $x_i=a$ 以及取另外 $n/2$ 的值为 $x_i=b$ 可以使 s_x^2 的值达到最大. 如果 n 是一个奇数，仍然在两个端点 a、b 重新选取所有的值，但是必须有一个端点比另一个端点多取一个观察值.

从上述讨论中可知如果要设计试验使得 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的均方误差都最小化，那么在给定的试验中值 $x_1,\cdots,x_n$ 应该正好有一半或接近 $n/2$ 的值等于某个尽可能大的数 c，而剩余的值等于 $-c$. 用这种方法，$\bar{x}_n$ 的值会恰好或近似等于 0，s_x^2 的值会尽可能的大.

最后，假设要估计线性组合 $\theta=c_0\beta_0+c_1\beta_1+c_*$，其中 $c_0\neq0$，且要设计试验使得 $\hat{\theta}$ 的均方误差最小化，即要使 $\mathrm{Var}(\hat{\theta})$ 最小化. 例如，Y 是一个对应于预测值 x 的将来的观察值，那么为了使 $\theta=\mathrm{E}(Y|x)$，可以令 $c_0=1,c_2=x,c_*=0$. 在例 10.2.2 中，我

们计算了当 $T=\hat{\theta}$ 时的 $\mathrm{Var}(T)$，它作为（10.2.11)式中两个非负项的和. 第二项是唯一依赖于 $x_1,\cdots,x_n$ 的取值的项，而且当且仅当 $\bar{x}_n=c_1/c_0$ 时，它的值等于0（它的最小可能取值）. 在该种情况下 $\mathrm{Var}(\hat{\theta})$ 取到其最小值 $c_0^2\sigma^2/n$.

616

实际上，一个有经验的数据分析员不会如刚才所述的最优设计那样，仅在一个单点或者仅仅在区间(a,b)的两个端点上取遍所有值 $x_1,\cdots,x_n$. 这是因为当所有 n 个观察值都仅取 X 的一个或二个值时，试验根本无法提供检验 Y 对 X 的回归是一个线性函数的假定的可能性. 为了检验该假定而不增加最小二乘估计量的均方差，$x_1,\cdots,x_n$中的许多值必须在 a、b 两个端点上选取，但是至少有一些值要在区间内的一些点上取到. 可以通过描画这些点来从直观上检验线性性，或用2次或更高次数的多项式来拟合.

10.2.6　小结

考虑如下统计模型，假定 $x_1,\cdots,x_n$ 已知，随机变量 $Y_1,\cdots,Y_n$ 是相互独立的，且 Y_i 服从均值为 $\beta_0+\beta_1x_i$ 和方差为 σ^2 的正态分布，其中 β_1、β_0、σ^2 都是未知参数. 这些都是简单线性回归模型的假设. 在该模型下，最小二乘估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的联合分布是一个二元正态分布，其中 $\hat{\beta}_i$ 的均值分别为 $\beta_i,i=1,2$. 其方差如(10.2.9)和(10.2.7)所给定. 其协方差如在(10.2.10)式中所给出. 如果考虑要预测与观察值 x 相对应的将来 Y 的值，可以用估计 $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1x$. 在该种情况下，$Y-\hat{Y}$ 服从均值为0和方差由(10.2.14)式给出的正态分布.

习题

1. 证明：σ^2 的均方差由（10.2.3)式给出.
2. 证明：$\mathrm{E}(\hat{\beta}_1)=\beta_1$.
3. 证明：$\mathrm{E}(\hat{\beta}_0)=\beta_0$.
4. 证明：$\mathrm{Var}(\hat{\beta}_0)$如(10.2.9)式中所给出.
5. 证明：$\mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1)$如(10.2.10)式中所给出.
6. 证明：在一个简单线性回归问题中，当 $\bar{x}_n=0$ 时，估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是独立的.
7. 考虑一个简单线性回归的问题，病人对新药 B 的反应 Y 与对标准药物 A 的反应 X 相关联. 假设可以得到如表10-1所示的十组观察数据.
   a. 确定 $\hat{\beta}_0$、$\hat{\beta}_1$ 和 $\hat{\sigma}^2$ 的极大似然估计.
   b. 确定 $\mathrm{Var}(\hat{\beta}_0)$和 $\mathrm{Var}(\hat{\beta}_1)$的值.
   c. 确定 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的相关系数的值.
8. 重新考虑习题7的情况，假定想要估计 $\theta=3\beta_0-2\beta_1+5$ 的值. 确定 θ 的一个无偏估计并计算它的均方差.

9. 重新考虑习题7中的条件，令 $\theta = 3\beta_0 + c_1\beta_1$，其中 c_1 为一个常数，确定 θ 的一个无偏估计 $\hat{\theta}$，求使 $\hat{\theta}$ 的均方差达到最小的 c_1 值.
10. 考虑习题7的情况，如果一个特殊的病人对药物 A 的反应值为 $x=2$，他对药物 B 反应的预测值应为多少，该预测的均方差为多少？
11. 考虑习题7中的条件，当病人对药物 A 的反应值 x 为多少时，他对药物 B 的反应的均方差会达到最小？

617

12. 考虑一个简单线性回归问题，一种特定类型的合金的柔韧度 Y 与生产它时的温度 X 相关. 假设在表10-3中给出了八对观察值. 求 $\hat{\beta}_0$、$\hat{\beta}_1$ 和 $\hat{\sigma}^2$ 的极大似然估计，并求出 $\text{Var}(\hat{\beta}_0)$ 和 $\text{Var}(\hat{\beta}_1)$ 的值.
13. 同习题12的情况，求出 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的相关系数.
14. 再次考虑习题12中的条件，假设要估计 $\theta = 5 - 4\beta_0 + \beta_1$ 的值. 找到 θ 的无偏估计量 $\hat{\theta}$，确定 $\hat{\theta}$ 的值和 $\hat{\theta}$ 的均方误差.
15. 再次考虑习题12的情况，令 $\theta = c_1\beta_1 - \beta_0$，其中 c_1 是一个常数. 求 θ 的一个无偏估计量 $\hat{\theta}$. 求当 c_1 取何值时 $\hat{\theta}$ 的均方误差会最小？
16. 同样考虑习题12的情况. 如果合金的一个样本是在温度 $x=3.25$ 时生产的，那么该合金样本的柔韧度的估计值是多少？该估计值的均方误差是多少？
17. 同样考虑习题12的情况. 温度 x 取何值时，该合金样本柔韧度的预测值的均方误差会最小.
18. Moore and McCabe(1999，第174页)报道了1970年和1980年几种海鲜食物的价格. 这些数据列在表10-6中. 如果我们试图根据1970年的价格来预计1980年的海鲜价格的话，那么可以用线性回归模型.
    a. 根据1970年的价格预计1980年的价格时找出最小二乘回归系数.
    b. 如果在1970年中添加一个售价为21.4的品种，你估计1980年它的售价会是多少？
    c. 1970年售价为21.4的品种，它在1980年预测价格的均方误差是多少？

表10-6 习题18中1970至1980年鱼的价格

1970	1980	1970	1980
13.1	27.3	26.7	80.1
15.3	42.4	47.5	150.7
25.8	38.7	6.6	20.3
1.8	4.5	94.7	189.7
4.9	23	61.1	131.3
55.4	166.3	135.6	404.2
39.3	109.7	47.6	149

19. 在十九世纪80年代，Francis Galton研究了生理特征的遗传. Galton发现高个子的儿子比平均身高要高，但是比他们的父亲要矮. 相似地，矮个子的儿子要比平均身高矮，但是比他们的父亲要高. 因此，儿子们的平均身高趋近于人口的平均身高，不管父亲比平均身高更高还是比平均身高矮. 从这些观测值，可以得出结论：身高的差距在一代代地减少，高

个和矮个都将消失，人口将“回归”到某个平均身高. 该结论是回归谬论的一个例子. 在本习题中将要证明此回归谬论会在二元正态分布中发生，甚至当两个坐标有相同的方差也会发生. 特别地，假设向量(X_1,X_2)服从相同均值μ、相同方差σ^2和正相关系数$\rho<1$的二元正态分布. 证明：对于每个x_1的值，$E(X_2|x_1)$要比x_1更趋近于μ. （尽管X_1与X_2有着同样的均值和方差，这仍然会发生.）

618

10.3 方差分析

在8.4节中，我们学习了比较两个正态分布均值的方法. 在本节中，我们要考虑比较两个或更多的正态分布均值所需要的实验. 由此方法所推导出的理论是完全建立在一般线性模型结果的基础上的.

10.3.1 单因子形式

本节要研究的主题就是*方差分析*，简称ANOVA. ANOVA问题实际上是多重回归问题，其设计矩阵Z有一个非常特殊的形式. 换言之，ANOVA的研究可以用一般线性模型框架来处理，该模型的基本假设不变：所得的观察值是独立的且服从正态分布；所得的观察值有相同的方差σ^2；每一个观察值的均值都可以用一个未知参数的线性组合来表示. ANOVA理论和方法论主要是在19世纪20年代由费希尔发展的.

首先通过考虑一个单因子设计问题来开始我们对ANOVA的研究. 在该问题中，假定随机样本来自于p个不同的正态分布，每个分布有着相同的方差σ^2，要在样本观察值的基础上比较这p个分布的均值. 在8.4节中，考虑了有两个总体$(p=2)$的情形，此处对任意p所得到的结果进一步推广了8.4节中的结果. 特别地，我们做如下的假设：对$i=1,\cdots,p$，随机变量Y_{i1}，…，Y_{in_i}形成一个由n_i个服从均值为μ_i和方差为σ^2的正态分布的观察值所组成的随机样本，此处$\mu_1,\cdots,\mu_p$和σ^2的值都是未知的.

在此问题中，样本容量$n_1,\cdots,n_p$不一定要都相等. 用$n=\sum_{i=1}^{p}n_i$表示p个样本里的观察值的总数，假设所有n个观察值都是独立的.

例10.3.1 热狗的卡路里 Moore and McCabe(1999)描述了由《消费者报道》(1986年6月第364页至367页）收集到的数据. 其中包含了63种品牌热狗的卡路里含量（见表10-7). 热狗有四种类型：牛肉、猪肉（不用说的)、禽肉、还有特色类.（特色类热狗含有例如奶酪或辣椒酱之类的馅料). 要知道不同类型的热狗的卡路里含量是否不同和不同到什么程度是很有意思的. 样本容量为$n_1=20$(牛肉)，$n_2=17$(猪肉)，$n_3=17$(禽肉)，$n_4=9$(特色类). 在此例中，μ_1代表牛肉类热狗的平均卡路里含量，μ_2代表猪肉类热狗的平均卡路里含量，μ_3代表禽肉类热狗的平均卡路里含量，μ_4代表特色类热狗的平均卡路里含量. 假设所有卡路里含量都是独立的且方差为σ^2的正态随机变量. 这些数据等我们学习完ANOVA方法以后再作分析.

665

表 10-7 四种类型热狗的卡路里含量

类型	卡路里含量
牛肉	186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135, 132
猪肉	173, 191, 182, 190, 172, 147, 146, 139, 175, 136, 179, 153, 107, 195, 135, 140, 138
禽肉	129, 132, 102, 106, 94, 102, 87, 99, 107, 113, 135, 142, 86, 143, 152, 146, 144
特色	155, 170, 114, 191, 162, 146, 140, 187, 180

◀

由刚才所做的假定，对于 $j=1,\cdots,n_i$ 和 $i=1,\cdots,p$，有 $\mathrm{E}(Y_{ij})=\mu_i$ 和 $\mathrm{Var}(Y_{ij})=\sigma^2$，由于每个观察值的数学期望 $\mathrm{E}(Y_{ij})$ 等于 p 个参数 $\mu_1,\cdots,\mu_p$ 中的一个，很明显每一个期望值可以看成是 $\mu_1,\cdots,\mu_p$ 的一个线性组合. 此外，我们可以把 n 个观测值 Y_{ij} 看作是一个可以写成以下形式的一个很长的 n 维向量 Y 中的元素，

$$\boldsymbol{Y}=\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{p1} \\ \vdots \\ Y_{pn_p} \end{bmatrix}. \tag{10.3.1}$$

因此这种单因子设计满足一般线性模型的条件. 为了使单因子设计看上去更像一般线性模型，可以定义参数 $\beta_i=\mu_{i+1}$，其中 $i=0,\cdots,p-1$. 那么在 $n\times p$ 阶设计矩阵 X 中，每一个总体有一对应列. 总体 1 的对应的列中有 n_1 个 1，接着是 $n_2+\cdots+n_p$ 个 0. 总体 2 中的对应列有 n_1 个 0，接着 n_2 个 1 和 $n_3+\cdots+n_p$ 个 0. 如此类推. 例如，用例 10.3.1 中热狗的数据，矩阵 Z 将变成

$$\boldsymbol{Z}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} 20\text{行} \\ \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} 17\text{行} \\ \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} 17\text{行} \\ \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} 9\text{行} \end{matrix} \tag{10.3.2}$$

666

由于参数 $\mu_1,\cdots,\mu_p$ 更为自然，所以在 ANOVA 的发展中我们将不再用一般线性模型符号.

对于 $i=1,\cdots,p$，令 $\overline{Y}_{i+}$ 代表第 i 个样本中 n_i 个观测值的样本均值. 即

$$\overline{Y}_{i+}=\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}. \tag{10.3.3}$$

可以证明(见习题1)$\overline{Y}_{i+}$是μ_i，$i=1,\cdots,p$的极大似然估计或最小二乘估计，且σ^2的极大似然估计为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i+})^2. \tag{10.3.4}$$

10.3.2 平方和的分解

在单因子设计下，我们通常对于检验从中抽取样本的这p个分布实际上是否相同这一假设感兴趣，也就是说要检验下面的假设：

$$H_0:\quad \mu_1=\cdots=\mu_p, \tag{10.3.5}$$
$$H_1:\quad 假设 H_0 是不真的.$$

例如，在例10.3.1中，(10.3.5)中的原假设H_0是：四种热狗的卡路里的均值是一样的，但没有指定该值是多少，备择假设H_1是其中至少有两个均值不同，但它没有指定哪些均值不同或这些均值相差多少.

在展开一个适当的检验过程之前，先要做一些预备性的代数运算. 首先，定义

$$\overline{Y}_{++}=\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}Y_{ij}=\frac{1}{n}\sum_{i=1}^{p}n_i\overline{Y}_{i+} \tag{10.3.6}$$

为所有n个观察值的均值. 再把平方和

$$S_{\text{Tot}}^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{++})^2.$$

分解为两个小一些的平方和，他们每一个将会和n个观察值中的某一种类型的方差相联系. 注意到如果我们相信所有的观察值来自于单个标准正态分布而不是来自于p个不同的正态分布，那么S_{Tot}^2/n将会是σ^2的极大似然估计. 这说明了可以把S_{Tot}^2解释为n个观察值的方差的一个总的度量. 在从S_{Tot}^2中分出了一个比较小的平方和可以度量不同样本间的方差，而另一个平方和可以度量在每个样本内的观察值之间的方差. 我们将要进行的假设(10.3.5)的检验中以两个方差度量的比值为基础，由于该原因，用*方差分析*这个名称来描述此类问题和其他相关问题.

667

如果我们只考虑样本i中的n_i个观察值，那么可以将那些值的平方和写成如下形式：

$$\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{++})^2=\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i+})^2+n_i(\overline{Y}_{i+}-\overline{Y}_{++})^2. \tag{10.3.7}$$

由定理7.3.1可知(10.3.7)式中右式的第一项服从一个自由度为n_i-1的χ^2分布，而且相对于$\overline{Y}_{i+}$是独立的. 既然$\overline{Y}_{++}$是$\overline{Y}_{i+},\cdots,\overline{Y}_{p+}$的一个函数，而这些项与等式(10.3.7)中右边的第一项是相互独立的，那么可知等式(10.3.7)中右边的两项是独立的.

如果把(10.3.7)式中的项关于i求和，就可以得到下列关系：

$$S_{\text{Tot}}^2=\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i+})^2+\sum_{i=1}^{p}n_i(\overline{Y}_{i+}-\overline{Y}_{++})^2. \tag{10.3.8}$$

令S_{Resid}^2表示(10.3.8)式右边的第一项求和，再令S_{Betw}^2表示另一个和. 那么，$S_{\text{Tot}}^2=$

$S^2_{Resid}+S^2_{Betw}$. 因为 p 个样本中所有观察值是相互独立的，S^2_{Resid}/σ^2 是每个都服从 χ^2 分布的独立的随机变量的和. 因此，S^2_{Resid}/σ^2 本身服从自由度为 $\sum_{i=1}^{p}(n_i-1)=n-p$ 的 χ^2 分布. 此外，S^2_{Resid} 和 S^2_{Betw} 也相互独立.

如先前所提到的，S^2_{Tot} 看作为所有这些观察值相对于总的均值的总方差. 类似地，可以将 S^2_{Resid} 看作为这些观察值关于它们特定的样本均值的总方差，或者是样本内剩余残差. 同时，可以将 S^2_{Betw} 看作是这些样本均值和总均值间的总方差，或者是样本均值间的方差. 所以，把总方差 S^2_{Tot} 分成了两个独立的分量 S^2_{Resid} 和 S^2_{Betw}，S^2_{Resid} 和 S^2_{Betw} 分别代表不同类型的方差. 这种分法经常被归纳在一个表里，如表 10-8 中所示的，该表格被称为是单个因子设计的方差分析表.

表 10-8 单因子设计的方差分析表的一般形式

方差来源	自由度	总平方和	均方
样本间	$p-1$	S^2_{Betw}	$S^2_{Betw}/(p-1)$
残差	$n-p$	S^2_{Resid}	$S^2_{Resid}/(n-p)$
总计	$n-1$	S^2_{Tot}	

在表 10-8 中“均方”这一列数是平方和除以自由度. 他们是用来检验假设(10.3.5). 若原假设(10.3.5)为真，那么在“样本间”和“总计”这两行中的自由度恰好是 χ^2 分布的随机变量的自由度. 通过建立对假设(10.3.5)的一个恰当的检验可以看到为什么该结论是正确的.

668

注： 剩余均方和回归场合下 σ^2 的无偏估计量是一样的. 本节开始的时候我们把单因子设计表示成数据向量为 Y 和设计阵 Z 的一个多重线性回归问题. 将 (10.3.4)式中 σ^2 的极大似然估计 $\hat{\sigma}^2$ 和表 10-8 中的残差均方作比较，可以看到两个仅在分母的常数上不一样. 当剩余均方为 $S^2_{Resid}/(n-p)$ 时，极大似然估计是 S^2_{Resid}/n. 记最后一个比例为 σ'^2，它是 σ^2 的无偏估计量（习题 5 中证明了这个事实）.

例 10.3.2 热狗中的卡路里 例 10.3.1 中的四个样本均值是

$$\overline{Y}_{1+}=156.85,\overline{Y}_{2+}=158.71,\overline{Y}_{3+}=118.76,\overline{Y}_{4+}=160.56.$$

总体均值是 $\overline{Y}_{++}=147.60$. 可以建立在表 10-9 中的方差分析形式. 在给出一个适当的检验统计量之后来检验假设(10.3.5). ◀

表 10-9 例题 10.3.2 的方差分析表

方差来源	自由度	总平方和	均方
样本间	3	19 454	6485
残差	59	32 995	559.2
总计	62	52 449	

10.3.3 检验假设

为了检验假设(10.3.5)，需要一种检验统计量，满足在 H_1 成立时比在 H_0 成立时的趋向要更大一些. 同时，还需要知道当 H_0 成立时该种检验统计量所服从的分布.

假定 H_0 为真，即设所有的观测值有相同的均值 μ. 在该种情况下，可以证明(参看习题 2) $S^2_{\text{Betw}}/\sigma^2$ 服从自由度为 $p-1$ 的 χ^2 分布. 我们已经看到 S^2_{Betw} 与 S^2_{Resid} 是相互独立的，且 $S^2_{\text{Resid}}/\sigma^2$ 服从自由度为 $n-p$ 的 χ^2 分布. 因此当 H_0 成立时，下面的随机变量 U^2 将服从自由度为 $p-1$ 和 $n-p$ 的 F 分布：

$$U^2 = \frac{S^2_{\text{Betw}}/(p-1)}{S^2_{\text{Resid}}/(n-p)}. \tag{10.3.9}$$

当原假设 H_0 不成立时，也就是至少有两个 μ_i 的值不同，那么 U^2 的分子的期望将大于它在 H_0 成立时的期望(见习题 11). 而无论 H_0 成立与否，U^2 的分母的分布总保持不变. 如果 $U^2 \geqslant F^{-1}_{p-1,N-p}(1-\alpha_0)$，那么该假设(10.3.5)在检验显著性水平 α_0 下将会拒绝 H_0，其中 $F^{-1}_{p-1,n-p}$ 是服从自由度为 $p-1$ 和 $n-p$ 的 F 分布的分位函数. 在本书后面给出了 F 分布的部分分位数表. 可以证明该检验也是检验水平为 α_0 的似然比检验过程(参看习题 12).

669
~
670

例 10.3.3　热狗中的卡路里　假如希望检验所有四种不同类型的热狗有相同均值的卡路里这一原假设，而其备择假设为至少有两种热狗有不同的均值. 如果原假设成立，(10.3.9)式中的统计量 U^2 服从自由度为 3 和 59 的 F 分布. U^2 的观察值为表 10-9 中样本均方与残差均方的比值，即 6485/559.2 = 11.60. 与该值相对应的尾部面积为 4.5×10^{-6}，因此原假设在大多数显著水平下被拒绝. ◀

检验的功效

如果(10.3.5)中的原假设不成立，那么等式(10.3.9)中的统计量 U^2 服从一个非中心的 F 分布. 为了得到功效函数的更多细节，可以参照更深的教材例如 Scheffé (1959，第二章). 此处我们将不进一步讨论方差分析检验的功效.

10.3.4　残差分析

因为单因子设计是普通线性模型的一个特殊情形，在进行单个因子方差分析计算时，我们要对普通线性模型做一些假设. 我们应该计算残差并画出它们的图来检验该假设是否合理. 当 $j=1,\cdots,n_i$ 和 $i=1,\cdots,p$ 时，残差的值是 $e_{ij}=Y_{ij}-\mu_i$.

例 10.3.4　热狗中的卡路里　图 10-8 是残差对种类变量“热狗类型”的散点图. 图 10-9 为残差对标准分位数的散点图. 在标准分位数散点图中的点是按照热狗类型来标注的. 在这些散点上出现一些干扰因素. 首先，有三个残差出现了相当大的负值. 其次，前三个样本中的每一个都包含两个相当明显的子集，一个含有较小残差而另一个则包含较大残差. 每个样本中的两个子集之间有一个间隔. 因而暗示着可能还有另外一个我们还没有讨论到的变量，而这个变量是不同于前两组的. 如果回到原始数据(消费报道的原始文章中)，会发现每一袋热狗的重量和每袋热狗的数量都被详细记录了. 两者之比是一只热狗的平均重量. 图 10-10 是热狗平均重量的残差散点图. 注意大多数大的残差来自于大的(重的)的热狗而残差小也趋向于来自于小的(轻的)热狗. 因此考虑把 Y 设置为每盎司的卡路里要比每只热狗的卡路里分析起来更好. ◀

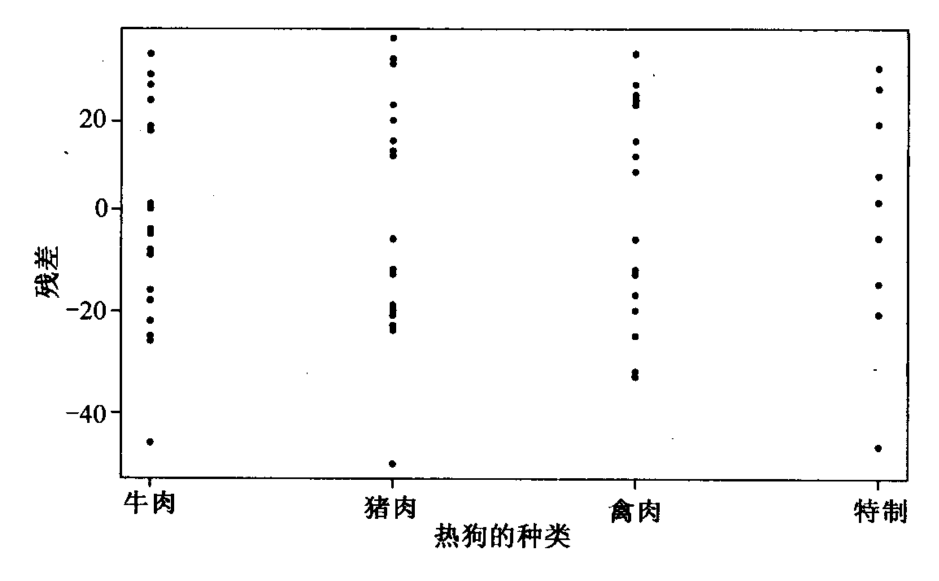


图 10-8 残差相对于热狗种类的图

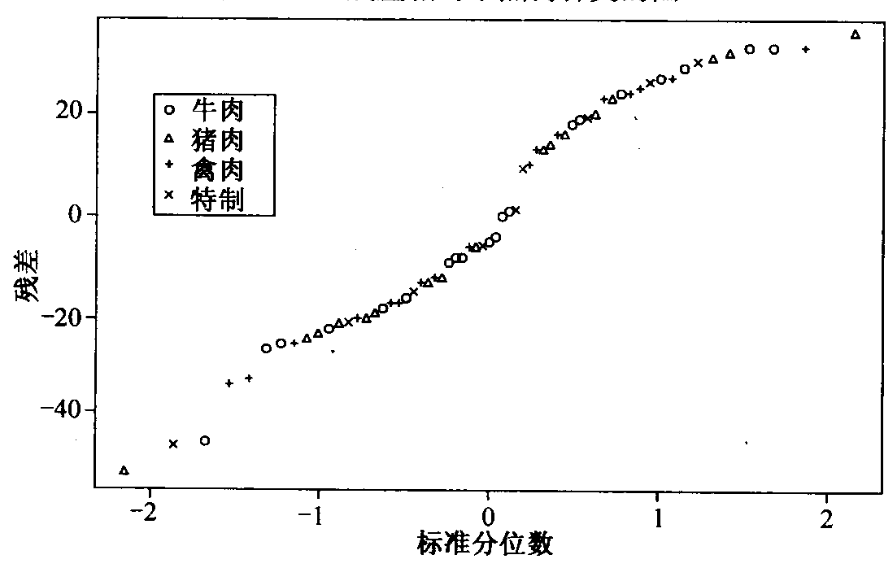


图 10-9 残差对标准分位数的散点图. 这些点是按照热狗种类描的

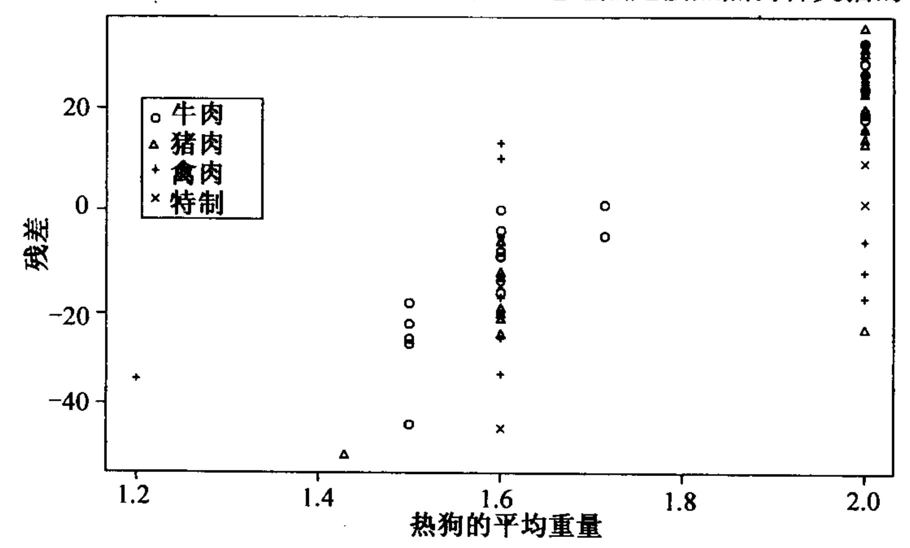


图 10-10 热狗平均重量的残差图. 按照热狗类型来描点

10.3.5　小结

可以把单因子设计看作是一个普通的线性模型，并且可以拟合该模型. 然而，在单因子设计中往往感兴趣的是(10.3.5)中的假设. 这些假设考虑的不仅仅是回归系数的一个线性组合，要去检验这些新假设，引进了方差分析和方差分析表. 如果 H_0 是真的，那么在(10.3.9)式中的统计检验量是服从自由度为 $p-1$ 和 $n-p$ 的 F 分布的 U^2. 如果 U^2 大于对应的 F 分布中 $1-\alpha_0$ 分位数，那么在检验水平 α_0 下拒绝 H_0.

671

习题

1. 在单因子设计中，通过证明对于 $i=1,\cdots,p$，向量 $(Z'Z)^{-1}Z'Y$ 的第 i 个分量是 $\overline{Y}_{i+}$ 来证明 $\overline{Y}_{i+}$ 是 μ_i 最小二乘估计.
2. 假设(10.3.5)中的 H_0 是真的，即所有的观察值有同一个均值 μ. 证明：$S^2_{\text{Betw}}/\sigma^2$ 服从自由度为 $p-1$ 的 χ^2 分布. 提示：令

$$X=\begin{pmatrix} n_1^{1/2}(\overline{Y}_{1+}-\mu)/\sigma^2 \\ \vdots \\ n_p^{1/2}(\overline{Y}_{p+}-\mu)/\sigma^2 \end{pmatrix},$$

然后用与7.3节中相同的方法来确定样本方差的分布. 无需证明可以直接使用下面的定理：设 $\boldsymbol{u}=((n_1/n)^{1/2},\cdots,(n_p/n)^{1/2})$，那么存在一个第一行为 $\boldsymbol{u}$ 的正交矩阵 $\boldsymbol{A}$.

3. 证明：

$$\sum_{i=1}^{p} n_i(\overline{Y}_{i+}-\overline{Y}_{++})^2=\sum_{i=1}^{p} n_i\overline{Y}_{i+}^2-n\overline{Y}_{++}^2.$$

4. 分析来自三个不同地区的一些牛奶场的牛奶样品，观察每个样品中放射性元素锶-90的含量. 假设样品来自第一个地区中的四个牛奶场、第二个地区的六个牛奶场和第三个地区的三个牛奶场. 结果以每升中的微微居里含量记录如下：

地区1：6.4，5.8，6.5，7.7，

地区2：7.1，9.9，11.2，10.5，6.5，8.8，

地区3：9.5，9.0，12.1.

a. 假设三个地区所有奶场的锶-90含量的方差是相同的，确定每个地区平均含量的极大似然估计以及共同方差的极大似然估计.

b. 检验三个地区的锶-90含量相同的假设.

5. 从四所大型高中的每个高年级班级中随机挑选10个学生，观察这40个学生在某次数学测验中的成绩. 假设每个学校的10个学生成绩的样本均值和样本方差如表10-10中所示. 检验假设：四所高年级在这次测验中的表现是没有差异的. 在进行这个检验时仔细讨论所做的假设.

672

表 10-10 习题 5 中的数据

学校	样本均值	样本方差
1	105.7	30.3
2	102.0	54.4
3	93.5	25.0
4	110.8	36.4

6. 假设样本容量为 n 的随机样本服从于均值为 μ、方差为 σ^2 的正态分布. 在观察样本前，将随机变量分成容量分别为 $n_1,\cdots,n_p$ 的 p 个组，其中 $n_i\geq 2$，对于 $i=1,\cdots,p$ 且 $\sum_{i=1}^{p}n_i=n$. 对 $i=1,\cdots,p$，令 Q_i 表示第 i 个组中 n_i 个观察值与样本均值偏差的平方和. 确定和 $Q_1+\cdots+Q_p$ 的分布以及比例 Q_1/Q_p 的分布.
7. 证明：如果 U 如(8.4.3)式所定义的，那么 U^2 等于(10.3.9)式所给出的表达式相同，利用此结果证明在 8.4 节中所给出的用于比较两个正态分布均值的 t 检验与本节中所给出的当 $p=2$ 时的单因子设计是相同的.
8. 证明：在单因子设计中，下面的统计量是 σ^2 的一个无偏估计：

$$\frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i+})^2.$$

9. 在单因子设计中证明：对于 i、i'和 j 的所有值，其中 $j=1,\cdots,n_i,i=1,\cdots,p$ 和 $i'=1,\cdots,p$. 下面的三个随机变量 W_1，W_2 和 W_3 之间两两互不相关.

$$W_1=Y_{ij}-\overline{Y}_{i+},W_2=\overline{Y}_{i'+}-\overline{Y}_{++},W_3=\overline{Y}_{++}.$$

10. 在 1973 年，Texaco 有限公司的总裁，写了一份关于空气和水污染的报告给美国参议院的一个委员会. 而该委员会关注的是噪音水平与汽车过滤器的关系. 他引用了表 10-11 中研究包括三种不同大小车辆噪音的数据.

表 10-11 习题 10 中的数据

车的大小	噪音值
小	810,820,820,835,835,835
中	840,840,840,845,855,850
大	785,790,785,760,760,770

a. 对于这些数据建立一个方差分析表.

b. 计算所有三种大小的车辆制造噪音的平均水平相同这一原假设的 p-值.

11. 如果(10.3.5)中的原假设 H_0 是错误的，证明：S^2_{Betw} 的期望值是 $(p-1)\sigma^2+\sum_{i=1}^{p}n_i(\mu_i-\overline{\mu})^2$，其中 $\overline{\mu}=\frac{1}{n}\sum_{i=1}^{p}n_i\mu_i$.
12. 证明：当 $U^2>F^{-1}_{p-1,n-p}(1-\alpha_0)$时，在单因子设计中假设(10.3.5)的水平 α_0 的似然比检验拒绝 H_0. 提示：第一部分 $\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2$ 形式上与(10.3.7)式相似. 在 S^2_{Tot}的公式中用一个常数，设为 μ，来代替 $\overline{Y}_{i++}$，把结果分解为与(10.3.8)式相似的形式，会出现一个额外的项.
13. 如果(10.3.5)的原假设是正确的. 证明：$S^2_{\text{Tot}}/\sigma^2$ 服从自由度为 $n-1$ 的χ^2 分布.

14. 单因子设计中一个常用的替代参数法如下：令 $\mu=\frac{1}{n}\sum_{i=1}^{p}n_i\mu_i$，并定义 $\alpha_i=\mu_i-\mu$. 这使得 $\mathrm{E}(Y_{ij})=\mu+\alpha_i$.

a. 证明：$\sum_{i=1}^{p}\alpha_i=0$.

b. 证明：α_i 的极大似然估计是 $\bar{Y}_{i+}-\bar{Y}_{++}$.

c. 证明：(10.3.5)中的原假设等价于 $\alpha_1=\cdots=\alpha_p=0$.

673

d. 证明：S_{Betw}^2的均值是 $(p-1)\ \sigma^2+\sum_{i=1}^{p}n_i\alpha_i^2$.

第11章 随机模拟

11.1 为什么随机模拟是有用的?

统计模拟常用于估计分布的特征，例如均值、分位数等以及其他不能得到显式表达的分布特征. 当使用模拟方法进行估计时，除了可以得到估计外，还可求得该估计的精度.

11.1.1 模拟的例子

模拟是一种技术，它能够帮助人们了解复杂系统是如何运作的，即使不能对它进行详细分析. 例如，工程师可通过模拟一个建筑工程附近的交通模型，来观察该工程受交通限制所产生的影响. 一个物理学家可以在不知气体分子运动机制的条件下模拟该气体分子的运动. 统计模拟则用来估计我们不能通过解析方法计算的模型的概率特征. 因为模拟方法在分析中引入了随机元素，它有时被称作“蒙特卡罗方法”，蒙特卡罗是欧洲著名的赌博中心.

例 11.1.1 样本中位数的均方误差(M. S. E) 考察一个容量为 n 的服从柯西分布的随机样本，分布的中心为 μ，且 μ 未知. 其中每个观察值的概率密度函数为

$$f(x)=\frac{1}{\pi}(1+[x-\mu]^2)^{-1},$$

未知参数 μ 是分布的中位数. 假设我们对于用样本中位数 M 作为 μ 的估计量的好坏感兴趣，特别地，要计算均方误差 $\mathrm{E}([M-\mu]^2)$. 如果能够产生一个容量为 n 的来自以 μ 为中心的柯西分布的随机变量组成的样本，那么就能计算样本中位数 M 和 $Y=(M-\mu)^2$，均方误差即为 $\theta=\mathrm{E}(Y)$. 如果能产生大量的 v 个与 Y 同分布的 $Y^{(1)},\cdots,Y^{(v)}$，则根据大数定理，$Z=\frac{1}{v}\sum_{i=1}^{v}Y^{(i)}$ 应该近似于 θ. 为此，我们可以产生 nv 个独立同分布且服从以 μ 为中心的柯西分布的随机变量. 然后把它们分成 v 个大小为 n 的集合并用每个集合中的所有 n 个变量来计算样本中位数 $M^{(i)}$ $(i=1,\cdots,v)$，然后计算 $Y^{(i)}=(M^{(i)}-\mu)^2$. 事实上有不少数据就是这么算出来的，这些数据包括了根据不同分布的随机样本计算得出的各种估计值的均方误差. ——◀

注: 不同模拟的记号. 用放在括号里的上标来区分同一随机变量的不同模拟值. 例如在例 11.1.1 中，我们用 $Y^{(i)}$ 代表 Y 的第 i 个模拟值. 类似地，可以模拟有下标的随机变量. 例如 $\mu_i^{(j)}$ 表示 μ_i 的第 j 个模拟值.

例11.1.1说明了许多统计模拟方法的主要特征. 如果我们想要计算的数可以表示成一些服从分布F的随机变量的期望，那么可以产生一个服从分布F的大样本并且取它们的平均数来得到该数. 但通常情况下，分布F本身非常复杂，例如例11.1.1中那样. 在这样的例子里，我们需要根据服从比较熟悉的分布的简单随机变量中构造出服从F分布的随机变量. 在例11.1.1中，均方误差是随机变量$Y=(M-\mu)^2$的平均值，其中M是以μ为中心，容量为n的柯西随机变量样本的样本中位数. 我们不可能一步就容易地模拟出服从Y分布的一个随机变量，但可以首先模拟n个柯西随机变量，然后找到它们的样本中位数M，最后计算出$Y=(M-\mu)^2$，其中Y的分布即为我们所需要的分布，在此之后我们多次重复模拟Y.

不是所有的统计模拟都用到随机变量的平均值.

例11.1.2 复杂分布的中位数 设X是参数μ未知的指数随机变量，假设μ服从概率密度函数为g的分布. 我们所关心的是X的中位数. X的边际概率密度函数为

$$f(x)=\int_0^\infty \mu e^{-\mu x}g(\mu)\,d\mu.$$

我们可能无法计算该积分，但假如可以产生一个由服从概率密度函数g的随机变量$\mu^{(1)},\cdots,\mu^{(v)}$所构成的大样本. 则对于每个$i=1,\cdots,v$，可以随机模拟出服从一个有参数$\mu^{(i)}$的指数分布的$X^{(i)}$. 随机变量$X^{(1)},\cdots,X^{(v)}$即为从概率密度函数$f$的分布中得到的一个随机样本，样本$X^{(1)},\cdots,X^{(v)}$的中位数则接近于概率密度函数$f$所构成的分布的中位数. ◀

例11.1.3 临床测试 考虑2.1节例2.1.3所描述的4组治疗组的例子. 对于$i=1,2,3,4$，令P_i为第i个治疗组的病人在治疗后不再旧病复发的概率，我们感兴趣的是P_i之间的差距达到特定值的程度. 假设P_i相互独立，其先验分布为参数为α_0和β_0的β分布. 从而P_i的后验分布亦相互独立，且为参数是α_0+x_i和$\beta_0+n_i-x_i$的β分布，其中n_i为第i组的病人个数，x_i是第i组病人不再发病的人数. 我们可以来模拟大量的v个服从上述β分布的(P_1,P_2,P_3,P_4)，这样就可以得到任何我们所想知道的关于(P_1,P_2,P_3,P_4)后验分布的问题的答案. 例如可以估计$\Pr(P_i>P_4)$，$i=1,2,3$，其中$i=4$代表未处理组. 该概率表明每种治疗方式比不治疗好的程度. 我们还可以通过计算向量(P_1,P_2,P_3,P_4)的样本中第i个分量大于第4个分量的比例来估计$\Pr(P_i>P_4)$. 我们也可以估计P_i是最大的概率，或者所有四个P_i相差都在ε之内的概率. ◀

[700]

例11.1.4 比较方差不相等的两个正态分布的均值 在第8章上我们讨论了两个方差未知且不相等的正态分布的均值是否相等的假设检验问题. 若在贝叶斯框架下使用模拟来做则会相对简单些. 此时参数为$\mu_x,\tau_x,\mu_y,\tau_y$，在参数给定的条件下，令$X_1,\cdots,X_m$独立同分布，其公共分布是均值为$\mu_x$、标准差为$\tau_x$的正态分布. 令$Y_1,\cdots,Y_n$

独立同分布(且与 X 是独立的)，是均值为 μ_y、标准差为 τ_y 的正态分布. 假设我们用自然共轭先验来定义 (μ_x,τ_x) 与 (μ_y,τ_y) 的先验分布，并设它们是独立的.（其实没有必要让 X 的参数与 Y 的参数独立,但这样可以使问题变得简单.）因为 X 数据及 X 参数均与 Y 数据及 Y 参数独立，所以可以分别计算两者参数的后验分布. 记 (μ_x,τ_x) 的后验分布的超参数为 α_{x1}、β_{x1}、μ_{x1} 和 λ_{x1}. 类似地，记 (μ_y,τ_y) 后验分布的超参数为 α_{y1}、β_{y1}、μ_{y1} 和 λ_{y1}. 为了对 $\mu_x-\mu_y$ 进行假设检验，需要知道 $\mu_x-\mu_y$ 的后验分布. 该分布不太容易分析. 如果由他们的联合后验分布来模拟一组参数矢量，对于每个样本向量可以计算 $\mu_x-\mu_y$，而且这些值将从 $\mu_x-\mu_y$ 的后验分布中形成一个样本. 更准确地说，令 v 是一个很大的数，那么对于每个 $i=1,\cdots,v$，我们想由联合后验分布来模拟 $(\mu_x^{(i)},\mu_y^{(i)},\tau_x^{(i)},\tau_y^{(i)})$. 为了达到此目的，需要模拟具有合适的后验分布的独立伽玛随机变量 $\tau_x^{(i)}$ 和 $\tau_y^{(i)}$. 在模拟之后，还可以模拟来自于均值为 μ_{x1} 和方差为 $1/(\lambda_{x1}\tau_x^{(i)})$ 的正态分布的 $\mu_x^{(i)}$. 类似地，还可以模拟来自于均值为 μ_{y1} 和方差为 $1/(\lambda_{y1}\tau_y^{(i)}$ 的正态分布的 $\mu_y^{(i)}$. 则对于 $i=1,\cdots,v$，$\mu_x^{(i)}-\mu_y^{(i)}$ 是来自于 $\mu_x-\mu_y$ 的后验分布的一个样本. 在我们讨论了对各类分布来模拟伪随机数的一些方法之后，在例 11.2.8 来举例说明该方法. ◀

例 11.1.4 中的模拟可以通过一个简单方式推广到具有不同方差的两个或多个正态分布的比较. 由于有多于两个均值要比较，所以就出现了要计算什么量来综合该比较. 即，不止像 $\mu_x-\mu_y$ 这样一个差可以获取三个或更多的均值的差异. 我们将在例 11.2.7 和例 11.3.7 中更为详细地讨论这种情形. 701

例 11.1.5 估计标准差 设 X 是一个随机变量，它的标准差 θ 是待估的一个重要的量. 假如我们无法算出 θ 的显式，但是可以模拟许多与 X 有相同分布的伪随机值 $X^{(1)},\cdots,X^{(v)}$. 那么，就可以计算出样本的标准差

$$S_v=\left(\frac{1}{v}\sum_{i=1}^{v}(X^{(i)}-\overline{X})^2\right)^{1/2}$$

作为 θ 的一个估计量. 因为 S_v 不是平均值，大数定理没有告诉我们它依概率收敛于 θ. 然而，如果令 $Y^{(i)}=X^{(i)2}$，就可以将 S_v 重写为 $(\overline{Y}-\overline{X}^2)^{1/2}$. 由这种形式，可以看到 $S_v=g(\overline{X},\overline{Y})$，其中 $g(x,y)=(y-x^2)^{1/2}$ 时. 注意 g 在满足 $y\geqslant x^2$ 的任意点上都是连续的. 大数定律告诉我们 $\overline{Y}$ 依概率收敛于 $\mathrm{E}(X^2)$ 以及 $\overline{X}$ 依概率收敛于 $\mathrm{E}(X)$. 因为 $\mathrm{E}(X^2)\geqslant\mathrm{E}(x)$，可以应用 4.7 节中的习题 16 来导出 S_v 依概率收敛于 $g(\mathrm{E}(X),\mathrm{E}(X^2))=\theta$. ◀

上面所有的例子都涉及了具有特定分布的大量的随机变量的生成. 该话题的某些讨论在第 3 章中出现过. 11.2 和 11.3 节将继续讨论具有特定分布的随机变量生成的方法. 11.4 节将重点介绍成功运用统计模拟的特殊的一类问题.

11.1.2 你指的是哪个均值?

模拟分析为已有的概率统计分析方法添加了一种新的概率分布和统计量的抽样分布的方法. 一个典型的统计分析应该包含了数据 $X_1,\cdots,X_n$ 的随机样本的一个概率模型. 该概率模型明确定义了每个 X_i 的分布, 以及此分布可能有的参数, 例如它的均值、中位数、方差和其他我们在估计和检验中感兴趣的测量值. 然后找一个统计值(数据的函数), 记为 Y. 这些函数可能包含我们想要估计的参数的样本的样式, 例如一个样本均值、样本中位数、样本方差, 诸如此类. 称 Y 的分布为它的抽样分布. 该抽样分布也可能有一个均值、中位数、方差和其他我们必须计算或处理的测量值. 迄今为止, 我们有了均值、中位数、方差和其他测量值的三个样式, 我们甚至还未开始讨论模拟.

可以用模拟分析来估计统计量 Y 的抽样分布中的参数 θ. 典型地, 模拟每个和 Y 有相同分布(样本的分布)的独立同分布伪随机变量 $Y^{(1)},\cdots,Y^{(v)}$. 然后计算出 $Y^{(1)},\cdots,Y^{(v)}$ 的一个综合的统计量 Z, 并且用 Z 去估计 θ. Z 本身可能是一个样本均值、样本中位数、样本方差, 或样本 $Y^{(1)},\cdots,Y^{(v)}$ 的其他度量. 称 Z 的分布为它的模拟分布或者是蒙特卡罗分布. 称模拟分布的特性, 如它的均值、中位数和方差为模拟均值、模拟中位数和模拟方差, 以便弄清楚我们已经到达了这个永远可以延伸的专业术语树的哪个水平. 下面的例子用来说明所有不同的水平.

702

例 11.1.6 五个或者更多的方差 设 $X_1,\cdots,X_n$ 是独立同分布随机变量, 且有分布函数为 F 的连续型分布. 令 ψ 表示 X_i 的方差. 假设我们决定用样本方差 $Y=\sum_{i=1}^{n}(X_i-\overline{X})^2/n$ 去估计 ψ. 在决定 Y 作为 ψ 的估计量是否合适时, 其中一部分要考虑它的方差 $\theta=\mathrm{Var}(Y)$. 即, θ 是 Y 的抽样分布的方差. 假设不能算出 θ 的显式, 但是假定可以容易地从 F 分布中模拟它. 对于 $j=1,\cdots,v,i=1,\cdots,n$, 我们也许要模拟 nv 个 $X_i^{(j)}$ 的值. 对于每个 j, 计算出样本 $X_1^{(j)},\cdots,X_n^{(j)}$ 的样本方差 $Y^{(j)}$. 即, $Y^{(j)}=\sum_{i=1}^{n}(X_i^{(j)}-\overline{X}^{(j)})^2/n$. $Y^{(j)}$ 的值都有与 Y 本身有相同的分布, 即 Y 的抽样分布. 因为我们对 $\mathrm{Var}(Y)$ 感兴趣, 所以可以计算出样本 $Y^{(1)},\cdots,Y^{(v)}$ 的样本方差 Z. 即, $Z=\sum_{i=1}^{v}(Y^{(i)}-\overline{Y})^2/v$. 然后用 Z 来估计 θ. 如果 Z 很大, 则意味着 Y 有很大的方差, 所以 Y 不是 ψ 的一个很好的估计量. 除非我们愿意收集更多的数据或者寻找一个更好的估计量, 否则就将被 ψ 的一个不好的估计量卡住了.

最后, 由于模拟的数量 v 或许还不足够多, Z 可能不是 θ 的一个好的估计量. 如果真是如此, 可以模拟更多的 $Y^{(j)}$ 值. 即, 可以增加模拟量 v 来获得 θ 的一个更好的模拟估计量(这不会使 Y 成为 ψ 的更好的估计量, 但是它将提供我们估计量好坏的一个更好的提示). 所以, 我们将试着去估计 Z 的方差(它的模拟方差). 如何准确地去做是随例题的变化而变化的,因此在这里我们将不做任何详细的叙述. 然

而，将在本节的稍后解释如何对最常用的模拟类型来估计 Z 的模拟方差.

方差的这个估计值必须在某处中止，且与 $\mathrm{Var}(Z)$ 一起结束. 即，我们将不会去评估 $\mathrm{Var}(Z)$ 的估计值有多好. 所有这些分布和估计的水平在表 11-1 中介绍. —◀

例 11.1.6 没有打算介绍任何模拟方法论. 它想要介绍不同水平的概率概念(例如方差)以及进入统计分析的模拟研究时的样本模式. 为了避免无望的混淆，区分正在讨论的是哪个方差或者是哪个样本方差是很重要的. 在本章中，我们将集中在模拟样本的特性上，特别是从模拟样本中计算出统计量的模拟分布. 然而，我们的例子必然会涉及在先前水平中出现的参数和统计量. 此外，模拟分布的分析将会使用与非模拟数据时使用的方法相同(中心极限定理，大数定律，δ 方法等等).

703

表 11-1 在典型模拟分析中概率分布、统计量和参数的水平

分布 (D) 或样本 (S)	参数 (P) 或统计量 (S)
(D) 人口分布 F	(P) 均值，方差，中位数等，ψ
(S) 从 F 来的样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$	(S) 基于 $\boldsymbol{X}$ 的 ψ 的估计值 Y，如样本均值，样本方差，样本中位数等.
(D) Y 的样本分布 G	(P) 均值，方差，中位数等，Y 的抽样分布 θ.
(S) 由 G 得到的模拟样本 $\boldsymbol{Y}=(Y^{(1)},\cdots,Y^{(v)})$	(S) 基于 $\boldsymbol{Y}$ 的 θ 的估计量 Z，如 $\boldsymbol{Y}$ 的样本均值，样本方差，样本中位数等.
(D) Z 的模拟分布 H	(P) 模拟分布的方差 (模拟方差).
(S) 模拟数据 (因例子而不同)	(S) 模拟方差的估计值 (依赖于特定的例题).

11.1.3 估计模拟结果的不确定性

在例题 11.1.6 中最后一步(在表格 11-1 中对于最后两行进行了总结)是对于每个模拟分析都很重要的一步. 即，应该总是试图去评估一个模拟中的不确定性. 通过模拟量的模拟方差非常容易评估该不确定性. 举例说明，在例题 11.1.1 中，令 $v=1000$ 和 $\theta=0$，可以产生 n 个柯西随机变量的 1000 个的样本，计算 $M^{(i)}$，第 i 个样本的中位数，并计算 $Y^{(i)}=(M^{(i)}-0)^2$ 的值. 则可以取这 1000 个 $Y^{(i)}$ 值的平均值. 可以重复该过程几次，但每次都不能得到同样的结果. 这要归于如下的事实，即使有像 1000 这样一个很大的 v，如 $Z=\frac{1}{v}\sum_{i=1}^{v}Y^{(i)}$ 这样的一个估计值仍然是一个具有正值方差的随机变量 (它的模拟方差). 模拟方差越小，就越能肯定估计值 Z 接近于我们想要估计的值. 但是，在能够评估不确定数量之前我们必须要估计或限制模拟的方差. 如何估计结果 Z 的模拟方差依赖于 Z 是否为模拟值的一个平均数，是否为一个或更多个平均值的一个平滑函数，或者是否为模拟值的样本分位数. 我们估计出的模拟方差的平方根被称为模拟标准差，而且它是 Z 的模拟标准差一个估计值. 模拟标准差能成为总结随机模拟不确定性的一个流行的办法有两个原因. 第一，它同所要测量的量具有相同的单位度量(不像模拟方差). 第二，模拟标准差对于说明模

拟估计值接近所要估计的参数的程度很有用. 在我们证明如何在几个常见的案例中

704

计算模拟标准差时将更详细地解释第二点.

例 11.1.7 一个样本分位数的模拟标准差 假设模拟分析的目的是为了估计某个随机变量 Y 的均值 θ. 模拟估计量 Z 通常是大量被模拟的值的平均数. 以下是一种直接估计一个平均数的模拟方差的方法. 假设要对 Y 进行大量的 v 次模拟，从而估计均值 θ. 即，假设对于一个很大的 v，我们模拟了互相独立的量 $Y^{(1)},\cdots,Y^{(v)}$. 如果还假定 θ 的估计值是 $Z=\frac{1}{v}\sum_{i=1}^{v}Y^{(i)}$，而且每个 $Y^{(i)}$ 都有均值 θ 和有限的方差 σ^2. 样本 $Y^{(1)},\cdots,Y^{(v)}$ 的样本标准差是样本方差的平方根，即为

$$\hat{\sigma}=\left(\frac{1}{v}\sum_{i=1}^{v}(Y^{(i)}-\bar{Y})^2\right)^{1/2}. \tag{11.1.1}$$

如果 v 是个很大的值，那么 $\hat{\sigma}$ 就很接近于 σ. 中心极限定理告诉我们 Z 应该近似服从于均值为 θ 和方差为 σ^2/v 的正态分布. 由于通常都不会知道 σ^2，所以应该用 $\hat{\sigma}^2$ 来估计其值. 从而使得 Z 的模拟方差的估计值就等于 $\hat{\sigma}^2/v$，且其模拟标准差为 $\hat{\sigma}/v^{1/2}$. ◀

例 11.1.8 关于另一个估计量的一个平滑函数的模拟标准差 有时候，在估计一个量 ψ 后，还希望估计它的一个平滑函数：$g(\psi)$. 例如，可能需要去估计某个均值的平方根或其对数. 或者，可能已经估计了一个方差 θ^2，而现在要求 θ 的一个估计. 通常地，假设通过模拟希望估计的参数是 $\theta=g(\psi)$，其中我们已经有了 ψ 的一个估计值 W. 进一步假设估计值 W 近以服从于均值为 ψ 和方差为 σ^2/v 的正态分布，其中 v 相对于 σ^2 相当大. 最后，假设在计算 W 时已经得到了 σ 的一个估计值 $\hat{\sigma}$. 例如，W 本身有可能就是均值为 ψ 和方差为 σ^2 的 v 个独立同分布的模拟随机变量 $Y^{(i)}$ 的平均值. 在该种情形下，(11.1.1)式将是 σ 的估计值. 令 $Z=g(W)$ 是 θ 的估计值. δ 的方法(见 5.7 节)告诉我们 Z 近似服从一个均值为 $\theta=g(\psi)$ 且方差为 $[g'(\psi)]^2\sigma^2/v$ 的正态分布. 例如，如果 $g(\psi)=\psi^{1/2}$，那么 $W^{1/2}$ 就服从均值为 $\psi^{1/2}$ 且方差为 $\sigma^2/[4\psi v]$ 的正态分布. 我们有了 σ 和 ψ 的估计，所以 Z 的模拟标准差为 $|g'(W)|\hat{\sigma}/v^{1/2}$. ◀

例 11.1.9 一个样本分位数的模拟标准误差 假设模拟分析的目的是去估计某个分布 G 的 p 分位数 θ_p. 典型地，我们模拟了数量很大的 v 个服从 G 分布的伪随机值 $Y^{(1)},\cdots,Y^{(v)}$，并用样本的 p 分位数作为估计值. 来自容量为 m 的一个大量随机样本中的样本分位数 p 近似服从以 θ_p 为均值和 $p(1-p)/[mg^2(\theta_p)]$ 为方差的正态分布，其中 g 是分布 G 的概率密度函数. 现在我们所关心的是该近似方差具有 σ^2/m 的形式，其中 $\sigma^2=p(1-p)/g^2(\theta_p)$ 是某个不依赖 m 的数. 假设模拟 k 个来自分布 G 且每

705

个样本容量为 m 的独立随机样本. 典型地，可通过选择原始模拟样本 $Y^{(1)},\cdots,Y^{(v)}$，

其容量为$v=km$，那么将v个模拟值分为k个容量为m的子样本. 计算这k个随机样本中的每个样本的p分位数，并称这些模拟的样本p分位数为$Z_1,\cdots,Z_k$. 为了对样本分位数使用近似正态分布，m必须很大. 接下来，计算$Z_1,\cdots,Z_k$的样本标准差

$$S=\left(\frac{1}{k}\sum_{i=1}^{k}(Z_i-\overline{Z})^2\right)^{1/2}. \tag{11.1.2}$$

其中$\overline{Z}$是k个样本中的p分位数的平均数. 如果把每个Z_i当作一个单独的模拟，那么S^2是Z_i的方差的一个估计值. 但是由前面可知Z_i的方差近似为σ^2/m. 因此S^2是σ^2/m的一个估计值. 换言之，σ的一个估计值为$\hat{\sigma}=m^{1/2}S$. 最后，把所有k个样本合并为单个容量为$v=km$的样本，且计算Z的样本p分位数，并把它作为θ_p的蒙特卡罗估计量. 如前所注意的，Z近似地服从均值为以θ_p和方差为σ^2/v的正态分布. 我们已构造了σ的一个估计量$\hat{\sigma}$，所以Z的模拟方差的估计值为$\hat{\sigma}^2/v=mS^2/v=S^2/k$，并且拟合标准误差为$S/k^{1/2}$. ◀

例 11.1.10 样本方差的模拟标准误差 假定模拟分析的目标是估计某个估计量Y的方差θ(例 11.1.6 就是建立在这种情况下的). 假定模拟$Y^{(1)},\cdots,Y^{(v)}$并用$Z=\frac{1}{v}\sum_{i=1}^{v}(Y^{(i)}-\overline{Y})^2$去估计$\theta$. 现在需要估计$Z$的模拟方差. 将$Z$重写为两个平均值的光滑函数，然后应用$\delta$方法的二维推广(见习题12)来估计模拟方差. 令$W^{(i)}=Y^{(i)2}$，可得$Z=\overline{W}-\overline{Y}^2$，其中$\overline{W}$是$W^{(1)},\cdots,W^{(v)}$的平均值. 现在$Z$是两个平均值的一个光滑函数，这时可应用习题12中所推导的二维δ方法(具体细节可见习题13). 习题13的结果提供了以下Z的渐近方差的近似. 首先，算出$W^{(1)},\cdots,W^{(v)}$的样本方差，并称其为V. 接下来计算出Y和W之间的样本协方差:

$$C=\frac{1}{v}\sum_{i=1}^{v}(Y^{(i)}-\overline{Y})(W^{(i)}-\overline{W}).$$

$\text{Var}(Z)$的估计就为

$$\widehat{\text{Var}(Z)}=\frac{1}{v}(4\overline{Y}^2Z-4\overline{Y}C+V). \tag{11.1.3}$$

而且，Z的模拟分布是均值为θ和方差为由等式(11.1.3)中给出的近似正态分布. 模拟标准误差是(11.1.3)的平方根. ◀

模拟的量是否足够?

令Z是基于v个模拟的某个参数θ的蒙特卡罗估计量. 由于现在可以估计出Z的模拟方差，所以可以回答Z接近于θ的程度的问题了. 而且还可以确定是否需要更多的模拟使Z足够得接近θ. 假定，如果至今所考虑的所有情形一样，Z近似服从均值为θ和方差为σ^2/v的正态分布，其中σ^2是不依赖于模拟容量的一个数. 对任意$\varepsilon>0$，

706

$$\Pr(|Z-\theta|\le\varepsilon)\approx2\Phi(\varepsilon v^{1/2}/\sigma)-1 \tag{11.1.4}$$

其中 Φ 是标准正态分布的分布函数，可以用这种逼近方法来帮助确定 Z 与 θ 有多接近. 在方程(11.1.4)式中 Z 的模拟标准差中，可以用 1 来替代 $v^{1/2}/\sigma$ 去逼近 $|Z-\theta|\leq\varepsilon$ 的概率. 我们还可以用(11.1.4)式来确定如果 v 不够大的话，还需做多少次模拟. 例如，假设想让(11.1.4)式的概率等于 γ. 那么应该令

$$v=\left[\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\frac{\sigma}{\varepsilon}\right]^2. \tag{11.1.5}$$

由于很难事先知道 σ，所以通常通过预先做容量为 v_0 的模拟并通过预先所做的这些模拟来计算 $\hat{\sigma}$.

例 11.1.11 样本中位数的均方误差 不难看出，不失一般性在例题(11.1.11)中可以令 $\mu=0$. 原因如下：令 $M^{(i)}$ 为 $X_1^{(i)},\cdots,X_n^{(i)}$ 的样本中位数，其中每一个 $X_j^{(i)}$ 都是以 μ 为中心的柯西随机变量. 则 $M^{(i)}-\mu$ 也是 $X_1^{(i)}-\mu,\cdots,X_n^{(i)}-\mu$ 的样本中位数，且每个 $X_j^{(i)}-\mu$ 都是以 0 为中心的柯西随机变量. 由于计算是建立在 $Y^{(i)}=(M^{(i)}-\mu)^2$ 的值的基础上的，其中 $i=1,\cdots,v$，无论是否有 $\mu=0$，都会得到相同的结果. 所以，令 $\mu=0$. 使得 $Y^{(i)}=M^{(i)2}$，且现在 σ^2 是 $M^{(i)2}$的方差. 假定要使 θ 的估计量 $Z=\overline{Y}$ 与 θ 的误差在 $\varepsilon=0.01$ 范围内的概率为 $\gamma=0.95$. 即，想要 $\Pr(|Z-\theta|\leq0.01)=0.95$. 由于 Z 是一个平均数，可以通过等式(11.1.11)计算 σ 的估计 $\hat{\sigma}$. 假设模拟了 $v_0=1000$ 个的来自柯西分布的容量 $n=20$ 的样本，计算 1000 个 $Y^{(i)}$ 的值，然后计算 $\hat{\sigma}=0.3892$. 根据等式(11.1.5)，用 0.3892 来代替 σ，可知需要 $v=[1.96\times0.3982/0.01]^2=5820$. 因此，需要大约 4820 个或更多的模拟次数. ◀

完成了等式(11.1.5)所建议的其余模拟之后，需要重新计算 $\hat{\sigma}$. 如果比开始估计的量要大很多，则需要再进行另外的模拟了.

例 11.1.12 复杂分布的中位数 在例 11.1.2 中，假定在概率分布函数 g 是参数为 3 和 1 的伽玛分布的概率密度函数. 假设想要使中位数的估计量与中位数的真实值的误差在 0.001 内的概率为 0.99. 我们首先做一个 $v_0=10\ 000$ 的初步模拟. 然后从参数为 3 和 1 的伽玛分布中模拟 $\mu^{(1)},\cdots,\mu^{(10\,000)}$. 对于每一个 i，模拟 $X^{(i)}$ 服从参数为 $\mu^{(i)}$ 的指数分布. 再将 $X^{(1)},\cdots,X^{(10\,000)}$ 分成 $k=20$ 个每个容量为 $m=500$ 的样本，然后计算 20 个样本中每一个样本的中位数 $Z_1,\cdots,Z_{20}$. 在进行了这些初始模拟之后，假定观察到等式(11.1.2)中 $S=0.015\ 97$. 这使得是 $\hat{\sigma}=0.3570$. 用这个值代替(11.1.5)式中的 σ，其中 $\gamma=0.99$ 和 $\varepsilon=0.001$，可以得到 $v=845\ 747.4$. 这意味着总共需要 845 748 次模拟以便在该模拟的结果中达到我们所需要的置信程度. 为了检验，总共模拟了 900 000 个值，并且计算出样本中位数 0.293 以及基于 $k=100$ 的每个样本容量为 $m=6200$ 的子样的新的 S^2 值. 新的 $\hat{\sigma}$ 值是 0.4529. 将 0.4529 替代等式(11.1.5)中的 σ，得到一个新的 $u=1\ 360\ 939$，这意味着还需要进行 40 939 次模拟. ◀

707

11.1.4 扩展研究：模拟实际过程

在许多科学领域中，实际的物理或社会变化过程都被构造为含有随机项的模型. 例如，股票价格常被构成为如例5.6.8中服从对数正态分布一样. 许多包含等待时间和服务过程通常被建模为泊松分布. 在本书前面所建立的简单的概率模型只是在实际过程中建立该类模型的基础. 在此，我们将给出可以用已知的分布来构造稍微复杂一些的模型的两个例子. 可以通过模拟来简化这些模型的分析.

例 11.1.13 期权定价 在例5.6.8中，我们介绍了Black和Scholes(1973)关于期权定价的公式. 在那个例子中，期权是关于要购买价格是 q 的股票，股票在时刻 u(未来时刻)的价值是一个服从对数分布的随机变量 S_u. 许多经济学家认为在例5.6.8中的 $\ln(S_u)$ 的标准差 σ 不应该被视为是一个已知常数. 例如，可以把 σ 视为一个概率密度函数为 $f(\sigma)$ 的随机变量. 为了精确起见，应该继续假设 $S_u=S_0e^{(r-\sigma^2/2)u+\sigma u^{1/2}z}$，但是现在要假设 Z 和 σ 都是随机变量. 为了方便起见，假设它们是独立的. 令 Z 服从标准正态分布，且令 $\tau=1/\sigma^2$ 服从参数为 α、β 均已知的伽玛分布. 参数 α、β 有可能可以根据结合历史数据和股票分析的专家意见来估计股票价格的方差而得到. 例如：它们有可能是对一个股票价格样本运用贝叶斯分析而得到的后验超参数. 易知对于所有的 σ 都有 $E(S_u|\sigma)=S_0e^{ru}$. 因此对期望的全概率定理(定理4.7.1)意味着 $E(S_u)=S_0e^{ru}$. 而这是风险中性所需要的. 在例5.6.8中所考虑的期权价格是随机变量 $e^{-ru}h(S_u)$ 的均值，其中

$$h(s)=\begin{cases} s-q & \text{若 } s>q, \\ 0 & \text{其他.} \end{cases}$$

(5.6.12)的Black-Scholes公式是已知 σ 条件下的 $e^{-ru}h(S_u)$ 的条件均值. 为了估计 $e^{-ru}h(S_u)$ 的边际均值，可以根据 σ 的分布来模拟大量的值 $\sigma^{(i)}(i=1,\cdots,v)$，将每一个 $\sigma^{(i)}$ 代入(5.6.12)，并将结果取平均.

作为一个例子，假设与例5.6.8结尾处取相同的数值，其中 $u=1$，$r=0.06$，$q=S_0$. 这次，假设 $1/\sigma^2$ 服从参数为2和0.0127的伽玛分布(这些值使得 $E(\sigma)=0.1$，但是 σ 具有相当大的可变性). 可以从该分布中对 σ 取 $v=1\ 000\ 000$ 个值并对每一个值计算(5.6.12)的值. 在模拟中平均值是 $0.0756S_0$，而模拟的标准差是 $1.814S_0\times10^{-5}$. 期权价格仅仅略高于当假设 σ 已知时的价格. 当 S_u 的分布更为复杂时，可以直接模拟 S_u 和估计 $h(S_u)$ 的均值. ◀

708

在下面的例子中每一个模拟都要许多的步骤，但是每一个步骤都相对简单. 在实际模拟过程中，将一些简单的步骤合并成一个复杂的步骤是十分普遍的.

例 11.1.14　有不耐烦顾客的服务排队　考虑一列队伍中，顾客的到达人数是一个速度为 λ/小时的泊松过程. 假设服务该列队伍只有一个服务人员. 到达该队伍的每一个顾客都计算队伍的长度 r(包括正在接受服务的顾客在内)并且以概率 P_r 离开，其中 $r=1,2,\cdots$,离开队伍的顾客不会再加入队伍. 假如进入队伍中的每一个顾客等待着直到他前面的那个顾客正在接受服务，才移动到队伍的最前面. 在到达队伍最前列后顾客接受服务的时间是一个参数为 μ 的指数型随机变量. 假设所有的服务时间和到达时间之间是相互独立的.

可以用模拟来了解该列队伍的行为. 例如，可以估计在开始营业后一个特定时刻 t 时队伍中的顾客的期望人数. 为了此目的，可以模拟多次，譬如 v 次队伍的实现. 对于每一个第 i 次实现，计算 t 时刻有多少顾客 $T^{(i)}$ 在队列中. 然后估计量是 $\frac{1}{v}\sum_{i=1}^{v}T^{(i)}$. 为了模拟单个实现，过程如下：模拟泊松过程中的到达的时间间隔 $X_1,X_2,\cdots$,模拟参数为 λ 的独立同分布的指数型随机变量. 令 $T_j=\sum_{i=1}^{j}X_i$ 是第 j 个顾客到达的时间. 当有一个 k 使得 $T_k>t$ 时就停止模拟，在时刻 t，只有最早的 $k-1$ 个顾客进入队伍. 对于每个 $j=1,\cdots,k-1$，模拟一个服务时间 Y_j 服从参数为 μ 的指数分布. 令 Z_j 表示第 j 个顾客到达队伍最前面的时间，并且令 W_j 表示第 j 个顾客离开队伍的时间. 例如：$Z_1=X_1$ 和 $W_1=X_1+Y_1$. 对于 $j>1$，第 j 个顾客先计算队伍的长度，然后决定是否离开. 如果顾客 j 到达时，顾客 i 仍然在队伍中($i<j$)，令 $U_{i,j}=1$，如果顾客 i 已经离开队伍，则令 $U_{i,j}=0$. 那么

$$U_{ij}=\begin{cases}1 & 若\ W_i\geqslant T_j,\\ 0 & 其他.\end{cases}$$

709

当第 j 个顾客到达时队伍中的顾客数为 $r=\sum_{i=1}^{j-1}U_{i,j}$. 那么我们就模拟一个服从参数为 p_r 的伯努利分布的随机变量 V_j. 如果 $V_j=1$，顾客 j 离开队伍，从而 $W_j=T_j$. 如果顾客 j 留在队伍中，则该顾客到达队伍最前面的时刻为

$$Z_j=\max\{T_j,W_1,\cdots,W_{j-1}\}.$$

即，第 j 个顾客要么到达后马上就到达队伍的最前面(如果正好没有人在被服务)，或者所有前 $j-1$ 个顾客都已离开，就紧接排上去. 而且，如果顾客 j 留在队伍中，则 $W_j=Z_j+Y_j$. 对于每个 $j=1,\cdots,k-1$，当且仅当 $W_j\geqslant t$ 时第 j 个顾客在 t 时刻会在队伍里.

给出一个数值例子，假设 $\lambda=2$，$\mu=1$，$t=3$，和 $p_r=1-1/r$，其中 $r\geqslant1$. 假设最初 $k=6$ 的模拟时间间隔为

$$0.215,\ 0.713,\ 1.44,\ 0.174,\ 0.342,\ 0.382.$$

这些时间的前 5 个总和为 2.884，但所有这 6 个的总和是 3.266. 所以，在时刻 $t=3$

时至多有 5 个顾客在队伍里. 假设前 5 个顾客的模拟服务时间为

$$0.251,\ 2.215,\ 2.855,\ 0.666,\ 2.505.$$

由于还不知道当每个顾客 j 到达时将有多少顾客在队伍里, 所以我们不能提前模拟 V_j 的值. 图 11-1 列出了一条将要描述的过程的模拟时间线. 从顾客 1 开始, 其值 $T_1 = Z_1 = 0.215$ 和 $W_1 = 0.215 + 0.251 = 0.466$. 对于顾客 2, $T_2 = T_1 + 0.713 = 0.928 > W_1$, 所以当顾客 2 到达时已没有人在队伍里, 且 $Z_2 = T_2 = 0.928$, 那么, $W_2 = Z_2 + 2.215 = 3.143$. 对于顾客 3, $T_3 = T_2 + 1.44 = 2.368 < W_2$, 所以 $r = 1$. 因为 $p_1 = 0$, 且顾客 3 留在队伍里, 因此不需要模拟 V_3. 从而有 $Z_3 = W_2 = 3.143$, 以及 $W_3 = Z_3 + 2.855 = 5.998$. 对于顾客 4, $T_4 = T_3 + 0.174 = 2.542$. 因为 $W_1 < T_4 < W_2$, W_3, 所以有 $r = 2$ 个顾客在队伍里. 然后用参数 $p_2 = 1/2$ 的伯努利分布来模拟 V_4. 假设模拟得 $V_4 = 1$, 即顾客 4 离开, 由此忽略第 4 个顾客的模拟服务时间, 这使得 $W_4 = T_4 = 2.542$. 对于顾客 5, $T_5 = T_4 + 0.342 = 2.884$, 此时顾客 2 和顾客 3 仍然在队伍里. 需要用含参数 $p_2 = 1/2$ 的伯努利分布来模拟 V_5. 假如 $V_5 = 0$, 所以顾客 5 留在队伍里. 于是, $Z_5 = W_3 = 5.988$ 以及 $W_5 = Z_5 + 2.505 = 8.393$. 最后, 对于 $j = 2,3,5$, 有 $W_j \geq 3$. 这意味着正如图 11-1 说明的一样, 在时刻 3 时有 3 个顾客在队伍里. 毫无疑问, 对于一个大的模拟可以通过计算机编程来计算这些结果.

710

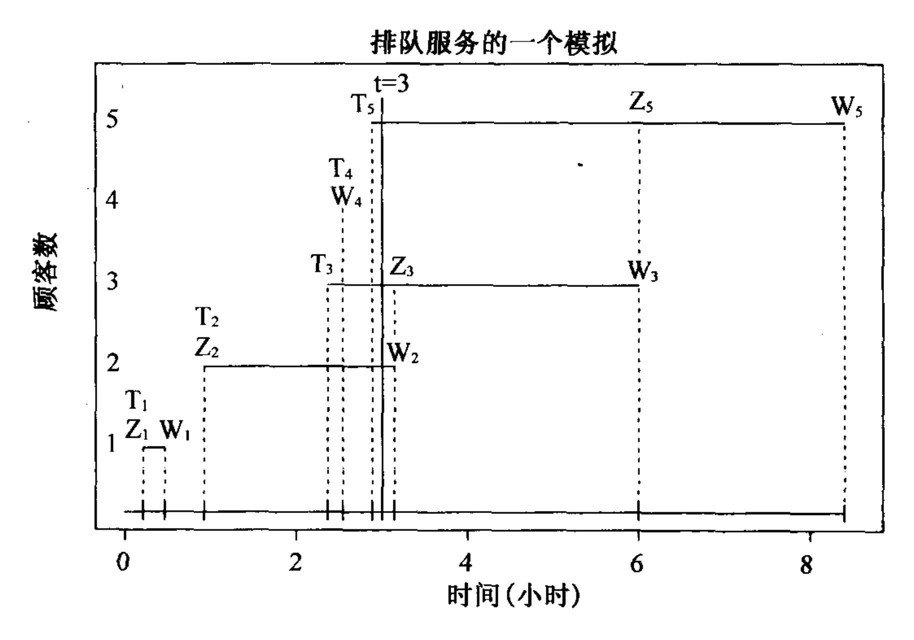


图 11-1 最底下的一条线是例 11.1.14 中的时间线. 每个顾客用一条水平线段来表示. 当 $t = 3$ 时与水平线重直的垂直线表示在时刻 $t = 3$ 时仍然在队伍中的那些顾客.

◀

11.1.5 小结

如果希望计算某个随机变量 Y 的期望值 θ, 但当不能进行必要的解析形式计算时, 可以使用模拟. 通常地, 要模拟和 Y 分布相同的一个大量的随机样本 $Y^{(1)}, \cdots,$

$Y^{(v)}$，然后计算样本均值 Z 作为估计量. 也可以用类似的方式来估计分布的分位数 θ_p. 如果 $Y^{(1)},\cdots,Y^{(v)}$ 是来自于分布的一个大样本，可以计算出样本的 p－分位数 Z. 如果能够计算出测量一个模拟估计量好坏程度的一些度量是一个好的想法. 一个普遍的度量是 Z 的模拟标准误差，即 Z 的模拟分布的标准差的估计. 当然，也可以通过进行足够大量的模拟来确保 Z 接近被估计参数的概率很高.

习题

1. (11.1.4)式是基于 Z 服从近似正态分布的假设. 有时候，正态的近似值不是很好. 在这种情况下，我们可以令

$$v=\frac{\sigma^2}{\varepsilon^2(1-\gamma)}. \tag{11.1.6}$$

为了更精确，令 Z 是 v 个具有均值为 μ 和方差为 σ^2 的独立随机变量的平均值. 证明：当 v 至少大于等于 (11.1.6)式中的值时，有 $\Pr(|Z-\mu|\leqslant c)\geqslant\gamma$. 提示：利用车贝雪夫不等式(4.7.2).

2. 在例 11.1.11 中，根据(11.1.6)式，需要 v 的值有多大?
3. 假设能够得到我们想要的足够多的独立同分布的标准正态随机变量. 令 X 表示服从均值为 2 和方差为 49 的正态分布的随机变量. 试描述一个利用模拟来估计 $\mathrm{E}(\ln(|X|+1))$ 的方法.
4. 利用伪随机数发生器来模拟一个有 15 个独立观察值的样本，其中有 13 个服从区间 $[-1,1]$上的均匀分布，而另外 2 个服从区间$[-10,10]$上的均匀分布. 利用所得到的 15 个值试计算：(a) 样本均值；(b) 对 $k=1,2,3,4$ 修正均值；(c) 样本中位数. 问这些估计量中哪个最接近于 0?
5. 将习题 4 重复 10 次，每次用不同的伪随机样本. 换言之，建立 10 个独立样本，每个样本有 15 个观测值，并且他们都满足习题 4 中的条件.
   a. 对每个样本，习题 4 中列出的哪个估计量最接近于 0?
   b. 对习题 4 中列出的每个估计量，试确定这 10 个样本中每个样本的估计量与 0 的距离的平方，并确定这 10 个平方距离的平均值. 所有估计量中哪个估计量距离 0 的平均平方距离最小?

711

6. 假设 X 和 Y 是相互独立的，且 X 服从参数为 3.5 和 2.7 的 β－分布，Y 服从参数为 1.8 和 4.2 的 β－分布. 我们对于 $X/(X+Y)$ 的均值感兴趣. 假定能够模拟你所希望的足够多的服从任意 β－分布的随机变量.
   a. 如果可以进行足够多的模拟，描述可以产生 $X/(X+Y)$ 均值的较好估计.
   b. 要使估计量与 $\mathrm{E}[X/(X+Y)]$的真实值相差不超过 0.01 的置信度为 98%. 描述如何确定模拟量的适当容量.
7. 假设你可以得到足够多的标准正态随机变量和在$[0,1]$区间内的均匀分布的随机变量. 希望通过进行一个模拟来得到位于“样本中位数”行和 $\varepsilon=0.05$ 列的数.
   a. 描述如何进行这样的模拟. 提示：设随机变量 X 和 U 是相互独立的，且 X 服从一个标准正态分布和 U 服从一个$[0,1]$区间上的均匀分布. 令 $0<\varepsilon<1$，则可以得到分布

$$Y=\begin{cases} X & 若\ U>\varepsilon, \\ 10X & 若\ U<\varepsilon. \end{cases}$$

b. 用计算机进行该模拟.

8. 考虑在习题 7 中所描述相同的情况. 这次，考虑在"对 $k=2$ 的截尾平均值"行和 $\varepsilon=0.1$ 所在列中的数.

   a. 描述如何进行模拟来生成这个数.

   b. 用计算机进行模拟.

9. 在例 11.1.12 中，事实上可以计算出 X_i 的分布的中位数 θ 的显式表达式. 计算真实的中位数，并观察模拟的值与真实值之间相差多远. 提示：利用随机变量的全概率公式(3.6.12)和(5.9.8)式来找出 X 的边际概率密度函数. 那么就很容易得到分布函数和分位数函数了.

10. 设 $X_1,\cdots,X_{21}$ 是独立同分布的，且都服从参数为 λ 的指数分布. 令 M 表示样本中位数. 希望把 M 的均方误差作为 X_i 分布的中位数的一个估计量.

    a. 确定 X_i 分布的中位数.

    b. 令 θ 是当 $\lambda=1$ 时样本中位数的均方误差. 证明：样本中位数的均方误差通常都等于 θ/λ^2.

    c. 描述估计 θ 的一个模拟方法.

11. 在例 11.1.4 中，有一个稍微简单一点的方法来模拟来自 $\mu_x-\mu_y$ 的后验分布的样本. 假设可以模拟足够多的相互独立的 t 个伪随机变量，且具有要求的任意自由度. 解释如何运用该 t 个随机变量来模拟来自 $\mu_x-\mu_y$ 的后验分布的一个样本.

12. 令 $(Y_1,W_1),\cdots,(Y_n,W_n)$ 是一个独立同分布的随机向量样本，且具有有限的协方差矩阵

$$\Sigma=\begin{pmatrix}\sigma_{yy} & \sigma_{yw}\\ \sigma_{yw} & \sigma_{ww}\end{pmatrix}.$$

令 $\overline{Y}$ 和 $\overline{W}$ 是样本平均值. 令函数 $g(y,w)$ 分别具有关于 y 和 w 的连续偏导数 g_1 和 g_2. 令 $Z=g(\overline{Y},\overline{W})$. 则 g 在 (y_0,w_0) 点附近的二维泰勒展开式为

$$g(y,w)=g(y_0,w_0)+g_1(y_0,w_0)(y-y_0)+g_2(y_0,w_0)(w-w_0),\qquad(11.1.7)$$

再加上一个在这里被忽略的误差项. 令(11.1.7)式中的 $(y,w)=(\overline{Y},\overline{W})$ 以及 $(y_0,w_0)=(\mathrm{E}(Y),\mathrm{E}(W))$. 由(11.1.7)式的近似水平，可以证明：

$$\mathrm{Var}(Z)=g_1(\mathrm{E}(Y),\mathrm{E}(W))^2\sigma_{yy}$$
$$+2g_1(\mathrm{E}(Y),\mathrm{E}(W))g_2(\mathrm{E}(Y),\mathrm{E}(W))\sigma_{yw}+g_2(\mathrm{E}(Y),\mathrm{E}(W))^2\sigma_{ww}.$$

提示：运用 4.6 节中得到的随机变量线性组合方差的公式.

13. 用习题 12 中的二维 δ 法来得到如 (11.1.3)式中所给出的样本方差的模拟方差的估计量. 提示：分别用 $\overline{Y}$ 和 $\overline{W}$ 代替 $\mathrm{E}(Y)$ 和 $\mathrm{E}(W)$，并用样本方差和样本协方差来代替 Σ.

14. 随机变量 Y 的分布的偏度是分布的非对称性的度量. 偏度的公式如下：令 Y 的均值为 μ 和标准差为 σ，则偏度为 $\mathrm{E}([Y-\mu]^3)/\sigma^3$. 假设有足够多的与 Y 同分布的伪随机变量，描述估计 Y 分布偏度的一个模拟方法.

712

15. 假设在未来时刻 u 的股票价格是一个随机变量 $S_u=S_0\mathrm{e}^{\alpha u+W_u}$，其中 S_0 表示现价，α 是一个常数，而且 W_u 是一个分布已知的随机变量. 假设能得到足够多的与 W_u 独立同分布的随

机变量. 假设 W_u 的 m. g. f. $\psi(t)$ 已知，且在一个包含 $t=1$ 的区间上有限.

a. 为了使 $E(S_u)=e^{ru}S_0$，α 应该为何值?

b. 我们希望对在时刻 u 以价格 q 买入一股这样的股票的期权定价. 描述一下如何用模拟来估计该定价.

16. 考虑一个队伍中顾客到达服从速率为 λ/小时的泊松过程. 假设为该队伍服务的有两个服务员. 每个到达队伍的顾客都会计算队伍的长度 r(包括任何正在被服务的顾客)，并决定以概率 p_r 离开，$r=2,3,\cdots$. 一个离开队伍的顾客就不会再进入队伍了. 而进入队伍的顾客要等到两名服务员中至少有一名服务员有空才能轮到. 如果两个服务员同时都有空，顾客以 0.5 的概率随机地在二者中选择一个，且与其他所有随机变量相互独立. 对于服务员 $i(i=1,2)$，从服务开始，为一位顾客服务的时间(以小时计算)是一个参数为 μ_i 的指数随机变量. 假设所有服务时间之间是相互独立的，且和顾客到达时间也是相互独立的. 描述应该如何模拟在特定的时刻 t 队伍中的顾客数(包括正在被服务的顾客).

11.2 特定分布的模拟

为了进行统计的模拟，必须从各种分布中得到伪随机数. 在本节中，我们将介绍对于某些分布进行模拟的一些方法.

大多数带统计功能的计算软件包能产生在区间[0,1]服从均匀分布的伪随机数. 在本节接下来的部分中，假设能产生服从[0,1]上均匀分布的独立同分布的随机变量(伪随机数)的任意大的样本. 通常地，需要服从其他分布的随机变量. 本节的目的就是回顾一些可将服从均匀分布的随机变量转换成服从其他分布的随机变量的一般方法.

11.2.1 概率积分变换

在第3章中，我们介绍了用概率积分变换把服从区间[0,1]上均匀分布的随机变量 X 变换成一个服从连续且严格单调递增的分布函数 G 的随机变量 Y. 在此方法中令 $Y=G^{-1}(X)$，当很容易计算 G^{-1} 时，该种方法就比较好用.

例 11.2.1 产生指数分布的伪随机变量 假设令 Y 服从参数为 λ 的指数分布，其中 λ 为常数. 则 Y 的分布函数为：

[713]

$$G(y)=\begin{cases}1-e^{-\lambda y} & 若\ y\geq 0,\\ 0 & 若\ y<0.\end{cases}$$

可以很简单地对它求反函数得：

$$G^{-1}(x)=-\ln(1-x)/\lambda, 若\ 0<x<1.$$

如果 X 是服从区间[0,1]上均匀分布的随机变量，那么 $-\ln(1-X)/\lambda$ 服从参数为 λ 的指数分布. ◀

11.2.2 特殊目的算法

这里有一些不易计算所需分布函数 G 的反函数的情形. 例如，若 G 是标准正态分布的分布函数，那么 G^{-1} 只能近似计算得到. 然而，有一个更好的方法，就是可将两个在区间[0,1]上服从均匀分布的独立的随机变量变换成两个标准正态分布的随机变量. 这种方法由 Box 和 Muller(1958)提出.

例 11.2.2 生成两个独立的标准正态随机变量 令 X_1, X_2 在区间[0,1]上独立且服从均匀分布. 则 (X_1,X_2) 的联合密度函数是

$$f(x_1,x_2)=1, \quad \text{当 } 0<x_1,x_2<1.$$

定义

$$Y_1=[-2\ln(X_1)]^{1/2}\sin(2\pi X_2),$$
$$Y_2=[-2\ln(X_1)]^{1/2}\cos(2\pi X_2).$$

求其反函数为：

$$X_1=\exp[-(Y_1^2+Y_2^2)/2],$$
$$X_2=\frac{1}{2\pi}\arctan(Y_1/Y_2).$$

用 3.9 节中的方法，可以计算出其雅可比行列式，即反函数的偏导数构成矩阵的行列式，这个矩阵为：

$$\begin{pmatrix} -y_1\exp[-(y_1^2+y_2^2)/2] & -y_2\exp[-(y_1^2+y_2^2)/2] \\ \dfrac{1}{2\pi y_2}\dfrac{1}{1+(y_1/y_2)^2} & -\dfrac{y_1}{2\pi y_2^2}\dfrac{1}{1+(y_1/y_2)^2} \end{pmatrix}.$$

该矩阵的行列式即为：$J=\exp[-(y_1^2+y_2^2)/2]/(2\pi)$. 则 (Y_1,Y_2) 的联合密度函数为

$$g(y_1,y_2)=f(\exp[(y_1^2+y_2^2)/2],\arctan(y_1/y_2)/2\pi))\,|J|$$
$$=\exp[-(y_1^2+y_2^2)/2]/(2\pi).$$

即为两个独立且标准正态的变量的联合密度函数. ◀

11.2.3 接受/拒绝法

对其他分布有许多特殊方法. 这里我们将给出一个被广泛使用的一般的方法. [714] 这种方法称为*接受/拒绝法*. 即令 f 为一个概率密度函数，并设想从该密度函数中抽取伪随机变量. 假定存在另一个概率密度函数 g 满足以下两种性质：

- 我们可由 g 知道如何模拟伪随机变量.
- 存在一个常数 k，使得对于所有 x，有 $kg(x)\geqslant f(x)$.

为了模拟一个概率密度函数为 f 的单个变量 Y 需要下面的步骤：

1. 模拟一个概率密度函数为 g 的伪随机变量 X 和一个与之独立的来自于[0,1]上均匀分布的伪随机变量 U.

2. 若

$$\frac{f(x)}{g(x)} \geqslant kU, \tag{11.2.1}$$

则令 $Y=X$，运算结束.

3. 如果(11.2.1)不成立，则舍弃 X 和 U，重新回到第一步.

如果不止需要一个 Y，则可根据需要重复尽可能多的该过程. 现在来证明每个 Y 的概率密度函数都是 f.

定理 11.2.1　*在接受/拒绝法中，Y 的概率密度函数是 f.*

证明： 我们知道 Y 的分布是给定满足(11.2.1)式的 X 的条件下的条件分布. 即，令 A 是(11.2.1)式成立时的事件，令 $h(x,u|A)$ 为给定 A 的条件下 (X,U) 的条件联合概率密度函数，那么 Y 的概率密度函数为 $\int h(x,u|A)\mathrm{d}u$. 这是因为 Y 是由满足(11.2.1)式条件的 X 构成的. 给定 A 的情况下，(X,U) 的概率密度函数是

$$h(x,u|A)=\frac{1}{\Pr(A)}\begin{cases} g(x) & \text{若 } f(x)/g(x)\geqslant ku, 0<u<1, \\ 0 & \text{其他.} \end{cases}$$

可以直接计算 $\Pr(A)$，即求 $U\leqslant f(X)/[kg(X)]$ 的概率

$$\Pr(A)=\int_{-\infty}^{+\infty}\int_0^{f(x)/[kg(x)]} g(x)\mathrm{d}u\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{1}{k}f(x)\mathrm{d}x=\frac{1}{k}.$$

所以

$$h(x,u|A)=k\begin{cases} g(x) & \text{若 } f(x)/g(x)\geqslant ku, 0<u<1, \\ 0 & \text{其他.} \end{cases}$$

固定 x，对这个函数关于 u 取积分，即可得到基于 x 的 Y 的概率密度函数

715

$$\int h(x,u|A)\mathrm{d}u=k\int_0^{f(x)/[kg(x)]} g(x)\mathrm{d}u=f(x).$$ ■

这里有一个使用接受/拒绝法的例子.

例 11.2.3　β-分布的模拟　假设要模拟一个服从参数为 1/2 和 1/2 的 β-分布的随机变量 Y，Y 的概率密度函数是

$$f(y)=\frac{1}{\pi}y^{-1/2}(1-y)^{-1/2}, 0<y<1.$$

注意：这个概率密度函数是无界的. 然而，很容易推出

$$f(y)\leqslant\frac{1}{\pi}(y^{-1/2}+(1-y)^{-1/2}), \tag{11.2.2}$$

对于所有的 $0<y<1$ 可以将(11.2.2)式的右边写成 $kg(y)$ 的形式，其中 $k=4/\pi$，

$$g(y)=\frac{1}{2}\left[\frac{1}{2y^{1/2}}+\frac{1}{2(1-y)^{1/2}}\right].$$

其中 g 是由两个概率密度函数 g_1，g_2 一半对一半混合而得出，

$$g_1(x)=\frac{1}{2x^{1/2}},\qquad 0<x<1,$$

$$g_2(x)=\frac{1}{2(1-x)^{1/2}},\qquad 0<x<1. \tag{11.2.3}$$

可以很容易地用概率积分变换方法来模拟服从这些分布的随机变量. 为了模拟来自概率密度函数为 g 的随机变量 X，先模拟三个服从[0,1]上均匀分布的独立的随机变量 U_1,U_2,U_3. 若 $U_1\leqslant 1/2$，则对于 U_2 用概率积分变换得到来自 g_1 的 X. 若 $U_1>1/2$，则对于 U_3 用概率积分变换得到来自 g_2 的 X. 若 $f(X)/g(X)\geqslant kU_3$，则 $Y=X$. 否则，重复此过程. ◀

当使用接受/拒绝法时，必须经常地拒绝模拟值并重新模拟. 由定理 11.2.1 的证明可知接受一个值的概率为 $\Pr(A)$，即 $1/k$. k 值越大，就越难接受. 在习题 5 中，你将证明首次接受模拟值之前模拟次数的期望值为 k.

接受/拒绝法一个常见的特例是模拟一个给定某事件条件下的随机变量. 例如，令 X 是一个来自于概率密度函数为 g 的随机变量，假定希望得到当给定 $X>2$ 时 X 的条件分布. 那么，给定 $X>2$ 的条件下，X 的概率密度函数为

$$f(x)=\begin{cases}kg(x) & \text{若 } x>2,\\ 0 & \text{若 } x\leqslant 2.\end{cases}$$

其中，$k=1/\int_2^\infty g(x)\,dx$. 由于对于所有 x，都有 $f(x)\leqslant kg(x)$，因此就可使用接受/拒绝法了. 事实上，由于 $f(x)/g(x)$ 只有两个值 k 和 0. 接受/拒绝法中不必模拟均匀分布变量 U. 我们甚至不必计算 k 的值，只要在 $X\leqslant 2$ 的时候拒绝就可以了，这里给出一个用这种算法去解决 10.4 节还没有解决的一个问题的方法.

716

例 11.2.4 计算两阶段检验的大小 在 10.4 节中，我们研究了如何分析来自于有放回的数据. 在那一节中介绍了一个两阶段检验方法. 首先检验(10.4.11)式的假设，若接受原假设，则继续检验(10.4.13)式的假设. 但不幸的是，不能计算给定第一个检验接受原假设的条件下，第二个检验条件大小. 即，不能得到(10.4.15)的显式值. 然而可以用模拟的方法来估计条件大小.

这两个检验是建立在由(10.4.12)式定义的 U_{AB}^2 和(10.4.16)式定义的 V_A^2 的基础上的，若 $U_{AB}^2\geqslant d$ 第一个检验拒绝由(10.4.11)式定义的原假设，其中 d 是近似的 F 分布的分位数. 若 $V_A^2\geqslant c$，则第二个检验拒绝它的原假设，其中 c 是需要被确定的. 随机变量 U_{AB}^2 和 V_A^2 都是不同的均方比率. 特别地，它们有相同的分母 $MS_{\text{Resid}}=S_{\text{Resid}}^2/[IJ(K-1)]$. 为了确定第二个检验的合适的临界值 c，需要得到给定 $U_{AB}^2<d$ 且两个原假设均为真的情况下 V_A^2 的条件分布，可以从下面的分布中抽样. 令交互均方为 $MS_{AB}=S_{\text{Int}}^2/[(I-1)(J-1)]$，因子A的均方为 $MS_A=S_A^2/(I-1)$，那么 $U_{AB}^2=MS_{AB}/$

MS_{Resid}，$V_A^2 = MS_A/MS_{\text{Resid}}$. 所有这些均方都是独立的，且当原假设为真时它们都来自于不同的伽玛分布. 大多数统计计算软件包都可以模拟来自于伽玛分布的随机变量. 因此，开始模拟多个三元组(MS_{AB}，MS_{Resid}，MS_A). 那么，对于每一个模拟出的三元组，计算 U_{AB}^2和 V_A^2. 若 $U_{AB}^2 \geq d$，舍去对应的 V_A^2，未被舍去的 V_A^2 就是来自于我们需要的条件分布的随机样本. ◀

11.2.4　产生其他随机变量的函数

通常对于一个特定的分布有不止一种的模拟方法. 例如，假设一个分布为某个其他随机变量的函数的分布(χ^2，t 和 F 分布就可以这么做). 此时，有一个直接的方法可以模拟所想要的分布. 首先，从所定义的分布中模拟随机变量，然后计算它的合适的函数.

例 11.2.5　模拟 β-分布的另一种方法　已知以下结论成立：如果 U 和 V 相互独立，并且 U 是服从参数为 α_1 和 β 的伽玛分布，V 是服从参数为 α_2 和 β 的伽玛分布，那么 $U/(U+V)$ 是服从参数为 α_1 和 α_2 的 β-分布. 因此，如果有一种可以模拟伽玛随机变量的方法，就可以模拟出 β 分布随机变量. 例 11.2.3 中 $\alpha_1 = \alpha_2 = 1/2$，令 $\beta = 1/2$，那么 U 和 V 均为服从参数为 1/2 和 1/2 的伽玛分布，即是自由度为 1 的 χ^2 分布. 如果模拟两个独立的标准正态随机变量 X_1，X_2（例如，可使用例 11.2.2 的方法)，那么 χ_1^2 与 χ_2^2 相互独立并且都是服从自由度为 1 的 χ^2 分布. 从而 $Y = X_1^2/(X_1^2 + X_2^2)$ 为服从参数为 1/2 和 1/2 的 β-分布. ◀

717

另举一个例子，为了模拟一个自由度为 10 的 χ^2 随机变量，一种方法可以模拟 10 个独立同分布的标准正态变量，再把它们的平方值相加. 另一种方法，可以模拟 5 个来自于参数为 1/2 的指数分布的随机变量，然后把它们相加.

例 11.2.6　产生二维正态随机变量的伪随机数　假设我们希望模拟一个具有(5.10.4)式中所给出的概率密度函数的二维正态随机变量，该概率密度函数可以用下面的随机变量的联合概率密度函数构造：

$$X_1 = \sigma_1 Z_1 + \mu_1,$$

$$X_2 = \sigma_2[\rho Z_1 + (1-\rho^2)^{1/2} Z_2] + \mu_2, \tag{11.2.4}$$

其中，Z_1 和 Z_2 为独立同分布的标准正态分布. 如果我们用例 11.2.2 的方法去生成服从标准正态分布的 Z_1 和 Z_2，那么可以用(11.2.4)式把它们转换成 X_1 和 X_2，使其服从我们想要得到的二元正态分布. ◀

大多数的统计计算软件包都有能力模拟来自于常用的（已被命名的）连续分布的随机变量. 本节中的方法事实上仅用于模拟一些不常见的分布或者统计软件包中没有的情况.

11.2.5 一些模拟常见分布的例子

例 11.2.7 一元配置的贝叶斯分析 我们对于一元配置使用与 10.3 节中所介绍的相同的统计模型来进行贝叶斯分析，其中模型参数有不合适的先验(可以用一个合适的先验，但是一些额外的计算会转移我们对于模拟问题的注意力)，令 $\tau = 1/\sigma^2$，参数 $(\mu_1,\cdots,\mu_p,\tau)$ 的常用的不合适先验有概率密度函数 $1/\tau$，后验联合概率密度函数与 $1/\tau$ 和似然函数的乘积成比例. 设观察数据为 y_{ij}，$j=1,\cdots,n_i$，$i=1,\cdots,p$，似然函数为

$$(2\pi)^{-n/2}\tau^{n/2}\exp\left(-\frac{\tau}{2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2\right),$$

其中 $n=n_1,\cdots,n_p$. 为了简化该似然函数，我们把指数部分重新改写为

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 = \sum_{i=1}^{p}n_i(\bar{y}_{i+}-\mu_i)^2 + S^2_{\text{Resid}},$$

其中 $\bar{y}_{i+}$ 是 $y_{i1},\cdots,y_{in_i}$ 的平均数，且残差平方和为

$$S^2_{\text{Resid}} = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2.$$

718

这样，后验概率密度函数 g 与

$$\tau^{p/2}\exp\left(-\frac{\tau}{2}\sum_{i=1}^{p}n_i(\bar{y}_{i+}-\mu_i)^2\right)\tau^{(n-p)/2-1}\exp\left(-\frac{\tau}{2}S^2_{\text{Resid}}\right)$$

成比例. 该表达式很容易被看作为服从参数为 $(n-p)/2$ 和 $S^2_{\text{Resid}}/2$ 的伽玛分布的变量 τ 的概率分布函数与 p 个分别服从均值为 $\bar{y}_{i+}$ 标准差为 $n_i\tau$，$i=1,\cdots,p$ 的正态分布的随机变量 $\mu_1,\cdots,\mu_p$ 的密度函数的乘积. 因此参数的后验联合分布如下：给定 τ 的条件下，μ_i 相互独立且服从均值为 $\bar{y}_{i+}$ 标准差为 $n_i\tau$ 的正态分布. τ 的边际分布是一个参数为 $(n-p)/2$ 和 $S^2_{\text{Resid}}/2$ 的伽玛分布.

如果我们要从后验分布中模拟参数的一个大样本，可以回答该问题，就是从数据中得到了什么. 为此，首先模拟大量的 τ，其值为 $\tau^{(1)},\cdots,\tau^{(v)}$，大多数统计计算程序允许使用者去模拟第一参数为任意值和第二参数为 1 的伽玛随机变量. 所以，我们可以模拟服从参数是 $(n-p)/2$ 和 1 的伽玛分布的变量 $T^{(1)},\cdots,T^{(v)}$，然后令 $\tau^{(l)}=2T^{(l)}/S^2_{\text{Resid}}$，$l=1,\cdots,v$，对于每 l 个独立的模拟值 $\mu_1^{(l)},\cdots,\mu_p^{(l)}$ 进行模拟，其中 $\mu_i^{(l)}$ 是均值为 $\bar{y}_{i+}$ 和方差为 $1/[n_i\tau^{(l)}]$ 的正态分布.

特别地，考虑例 10.3.1 中的热狗数据．首先如上所描述的模拟 $v=60\ 000$ 组. 现在，可以提出问题，即均值之间的差距程度如何. 这里有一些方法，可以计算出对于任给的 $c>0$，$|\mu_i-\mu_j|>c$ 的概率；也可以计算出对于任给的 $c>0$，至少有一

个$|\mu_i-\mu_j|>c$的概率；可以算出$\max_{i,j}|\mu_i-\mu_j|$，$\min_{i,j}|\mu_i-\mu_j|$，或者$|\mu_i-\mu_j|$的均值的分位数. 例如，60 000 个模拟值中99%至少有一个$|\mu_i^{(l)}-\mu_j^{(l)}|>27.94$. $\max_{i,j}|\mu_i-\mu_j|$的99%的分位数的估计的模拟标准差是0.117(对于本例接下来的部分,我们将只给出模拟估计,而不给出它们的模拟标准差). 模似值大约1/2都满足$|\mu_i-\mu_j|>2.379$. 在有99%的模拟值中的平均值至少为14.59. 究竟27.94、14.59还是2.379表示大的差距取决于我们需要对热狗做什么样的决策. 总结所有这些计算的一个有效的方法是画出这六个$|\mu_i-\mu_j|$差值的最大、最小和平均值的样本分布函数图. 本例中的图形见图11-2. 如果只关心这四种热狗之间是否存在差异，有一个方法就是检查图11-2中的"最大值"曲线(你能解释为什么可以这么做吗?).

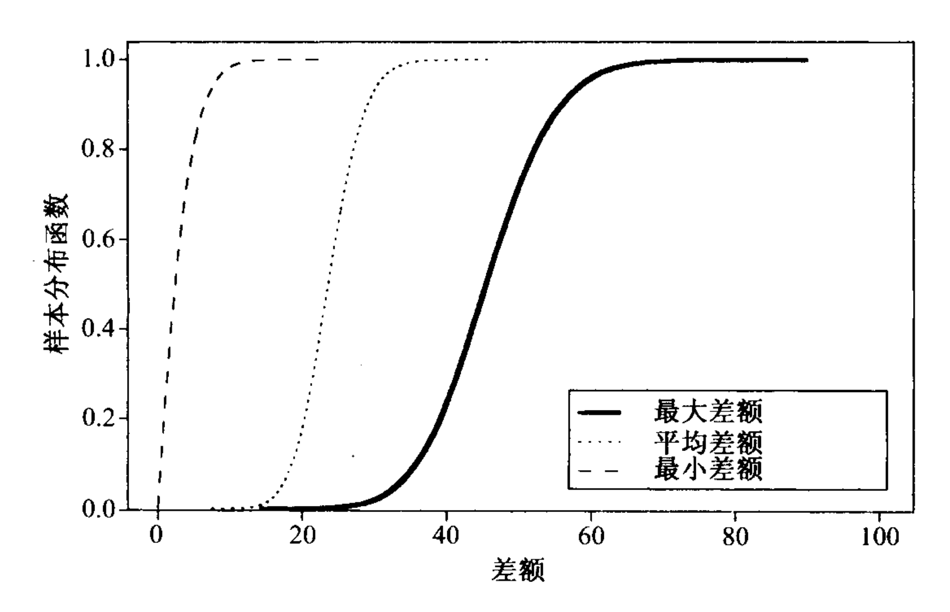


图11-2　例题11.2.7中六个$|\mu_i-\mu_j|$差额的最大、平均和最小值的样本分布函数

我们也可以试着去解决第十章中方差分析框架下的一些非常困难的问题. 例如，我们可能会问每个μ_i是四个值中的最大值或最小值的概率. 对于每个i，令N_i表示第j次模拟中，$\mu_i^{(j)}$是$\mu_1^{(j)},\cdots,\mu_4^{(j)}$中最小一个的个数. 再令$M_i$表示第$j$次模拟中，$\mu_i^{(j)}$是这四个均值中最大值的个数. 那么$N_i/60\ 000$就是$\mu_i$为最小均值的概率的模拟估计，$M_i/60\ 000$就是$\mu_i$为最大均值的概率的模拟估计. 结果见表11-2. 我们可以看出μ_3几乎一直是最小的，而μ_4有将近50%的可能性有最大的. ———◀

719

表11-2　例11.2.7中，每个μ_i是最大值和最小值的后验概率

类型	i	Pr(μ_i 最大 | y)	Pr(μ_i 最小 | y)
牛肉	1	0.1966	0
肉	2	0.3211	0
家禽	3	0	1
特产	4	0.4823	0

例 11.2.8 比较各种铜矿石 我们将利用例 8.4.2 中有关铜矿石的数据来说明例子 11.1.4 的方法，假设所有参数的先验分布都是不适当的. 观察到的数据包括一个容量为 8 和另一个容量为 10 的样本，并且 $\overline{X}=2.6$，$\sum_{i=1}^{8}(X_i-\overline{X})^2=0.32$，$\overline{Y}=2.3$，$\sum_{i=1}^{10}(Y_i-\overline{Y})^2=0.22$. 从而后验分布有超参数：$\mu_{x1}=2.6$，$\lambda_{x1}=8$，$\alpha_{x1}=3.5$，$\beta_{x1}=0.16$，$\mu_{y1}=1.15$，$\lambda_{y1}=10$，$\alpha_{y1}=4.5$，$\beta_{y1}=0.11$. τ_x 和 τ_y 的后验分布分别是参数为 3.5、0.16 和 4.5、0.11 的伽玛分布. 可以分别从这两个分布中很容易地抽取 10 000 个伪随机数. 对于每个模拟值 τ_x，从均值为 2.6 且方差为 $1/(8\tau_x)$ 的正态分布中抽得一个 μ_x，对于每个模拟值 τ_y，从均值为 2.3 且方差为 $1/(10\tau_y)$ 的正态分布中抽得一个 μ_y. 图 11-3 描绘了 10 000 个模拟值 $\mu_x-\mu_y$ 的直方图和 $\mu_x-\mu_y$ 的样本分布函数. 图形显示 $\mu_x-\mu_y$ 几乎总是为正. 事实上，有超过 99% 部分为正. $|\mu_x-\mu_y|<0.5$ 的概率非常大，因此若在该问题上 0.5 不算是一个大的差距，则认为 μ_x 与 μ_y 十分接近. 反之，若 0.1 是一个大的差距，则认为 μ_x 与 μ_y 相差较大. ◀

720

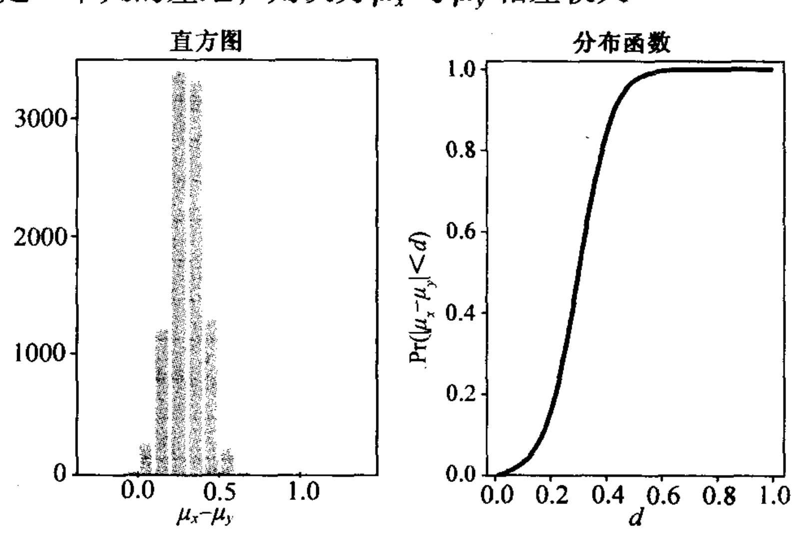


图 11-3 例题 11.2.8 中模拟的 $\mu_x-\mu_y$ 的直方图和 $|\mu_x-\mu_y|$ 的后验分布函数

如果只是关注例 11.2.8 中 $\mu_x-\mu_y$ 的分布，那么可以直接模拟 μ_x 与 μ_y，而不用先模拟 τ_x 和 τ_y. 由于在此例中，μ_x 与 μ_y 独立，我们便可以分别从它们各自的边际分布中来模拟.

例 11.2.9 *t* 检验的功效 在 8.3 节我们举例说明了如何基于非中心对称的 t 分布函数计算 t 检验的功效函数. 不是所有的统计软件包计算非中心对称的 t 的概率. 我们可以用模拟的方法来估计这些值. 令 Y 服从自由度为 m、非中心参数为 ψ 的 t 分布. 那么 Y 的分布与 $X_1/(X_2/m)^{1/2}$ 的分布相同，其中 X_1 和 X_2 相互独立. X_1 服从均值为 ψ、方差为 1 的正态分布. X_2 服从自由度为 m 的 χ^2 分布. 估计 Y 的分布函数的一个简单的方法就是模拟大量的 (X_1,X_2) 数对，然后计算 $X_1/(X_2/m)^{1/2}$ 的样本分布函数值. ◀

一个样本分布函数的模拟标准误差

在例11.2.7和例11.2.8中，我们绘制了模拟数据函数的样本分布函数. 这些函数中没有用到模拟标准误差. 可以对样本分布函数的每个值计算模拟标准误差，但是有一个更简单的方法可以概括样本分布函数的不确定性. 为了概括模拟的结果，令 $Y^{(i)}$，$(i=1,\cdots,v)$ 为被模拟出的具有分布函数为 G 的独立同分布样本. 令 G_v 为样本分布函数. 对于每个实数 x，$G_v(x)$ 为模拟样本中小于或等于 x 的比例. 即，$G_v(x)$ 为使得 $Y^{(i)}\leqslant x$ 的数值 i 的 $1/v$ 倍. 由相关知识可知如果 v 很大，那么，

721

$$\Pr(|G_v(x)-G(x)|\leqslant t/v^{1/2},\text{对于所有 } x)\approx H(t),$$

其中 H 为表11-3中的函数.

表11-3 H 的概率密度

t	$H(t)$	t	$H(t)$
0.30	0.0000	1.20	0.8878
0.35	0.0003	1.25	0.9121
0.40	0.0028	1.30	0.9319
0.45	0.0126	1.35	0.9478
0.50	0.0361	1.40	0.9603
0.55	0.0772	1.45	0.9702
0.60	0.1357	1.50	0.9778
0.65	0.2080	1.60	0.9880
0.70	0.2888	1.70	0.9938
0.75	0.3728	1.80	0.9969
0.80	0.4559	1.90	0.9985
0.85	0.5347	2.00	0.9993
0.90	0.6073	2.10	0.9997
0.95	0.6725	2.20	0.9999
1.00	0.7300	2.30	0.9999
1.05	0.7798	2.40	1.0000
1.10	0.8223	2.50	1.0000
1.15	0.8580		

特别地，$t=2$ 时，$H(t)=0.9993$. 所以可以断言（至少是近似地），$|G_v(x)-G(x)|\leqslant 2/v^{1/2}$ 对于所有的 x 以0.9993的概率同时成立. 在例11.2.7中，我们有 $v=60\ 000$，因此图11-2中的每个曲线应该以0.9993为概率被精确到0.008. 事实上，所有3条曲线同时被精确到0.008的概率为0.9979（请在习题14中证明）.

11.2.6 离散型随机变量的模拟

到现在为止，在本章中的所有例题，都只是涉及连续分布随机变量的模拟. 偶尔，会需要模拟离散分布的随机变量. 模拟离散型随机变量的算法是存在的，这里我们将介绍一些方法.

例 11.2.10 伯努利随机变量的模拟 模拟一个以 p 为参数的伯努利随机变量 X 的伪随机值是很简单的. 以一个在区间 $[0,1]$ 上服从均匀分布的 U 开始, 如果 $U\leqslant p$, 令 $X=1$, 否则, 令 $X=0$. 由于 $\Pr(U\leqslant p)=p$, X 就有一个正确的分布. 可以用该方法来模拟任何一个只有两个值支撑点的分布. 如果,

$$f(x)=\begin{cases}p & 若\ x=t_1,\\ 1-p & 若\ x=t_2,\\ 0 & 其他.\end{cases}$$

则当 $U\leqslant p$, $X=t_1$, 否则, $X=t_2$. ◀

例 11.2.11 模拟离散均匀随机变量 假设希望模拟来自于概率函数为下式的分布的伪随机变量,

$$f(x)=\begin{cases}\dfrac{1}{n} & 若\ x\in\{t_1,\cdots,t_n\},\\ 0 & 其他.\end{cases} \tag{11.2.5}$$

在整数 $1,\cdots,n$ 上的均匀分布是这种分布的一个例子. 下面给出模拟一个以(11.2.5)式为概率函数的随机变量的简单方法. 令 U 在 $[0,1]$ 上服从一个均匀分布, 并令 Z 为小于或等于 $nU+1$ 的最大的整数. 易知, Z 以相同的概率取值 $1,\cdots,n$, 从而 $X=t_Z$ 有(11.2.5)这样的概率密度函数. ◀

在例 11.2.11 中描述的此方法并不适用于更一般的离散分布. 然而, 例 11.2.11 中的方法对 1.5 节描述的自助法中的模拟是有用的.

722

对于一般的离散分布, 有一种概率积分变换的类似物. 假设关于点 $t_1<\cdots<t_n$ 的离散分布, 分布函数为

$$F(x)=\begin{cases}0 & 若\ x<t_1,\\ q_i & 若\ t_i\leqslant x<t_{i+1}, i=1,\cdots,n-1,\\ 1 & 若\ x\geqslant t_n.\end{cases} \tag{11.2.6}$$

下面是分位数函数:

$$F^{-1}(x)=\begin{cases}t_1 & 若\ 0<p\leqslant q_1,\\ t_{i+1} & 若\ q_i\leqslant p<q_{i+1}, i=1,\cdots,n-2,\\ t_n & 若\ q_{n-1}<p<1.\end{cases} \tag{11.2.7}$$

可以证明(见习题 13), 如果 U 服从在区间 $[0,1]$ 上的均匀分布, 那么 $F^{-1}(U)$ 有(11.2.6)式中的分布函数. 它给出了一个直观但并不有效的模拟任意的离散分布的方法. 注意将 n 限制为有限事实上是不必要的. 即使分布有无限多个可能的值, F^{-1} 也可以被(11.2.7)定义, 只要用 ∞ 代替 $n-2$, 并移去最后一段即可.

例 11.2.12 几何分布随机变量的模拟 假设要模拟来自于参数为 p 的几何分布的伪随机变量 X. 即在(11.2.7)式的注释中，$i=1,2,\cdots$时 $t_i=i-1$，$q_i=1-(1-p)^i$. 使用概率积分变换，可以先模拟服从区间 $[0,1]$ 上均匀分布的 U，然后比较 U 及 q_i，$i=1,2,\cdots$，直到首次出现 $q_i<U$，即令 $X=i$. 在本例中，由于 q_i 有简单的表达式，可以避免一系列的比较，使得 $q_i<U$ 的第一个 i 为一个严格小于 $\ln(1-U)/\ln(1-p)$ 的最大整数. ◀

但这种概率积分变换对于可能取值很大而 q_i 又没有一个简单表达式的离散分布而言是非常低效的. Walker(1974)，Kronmal 和 Peterson(1979)提出了一个更有效称作“Alias”的方法. 该种方法的操作如下：令 f 为我们想要从中模拟随机变量 X 的概率函数. 假设仅对于 x 的 n 个不同的取值有 $f(x)>0$. 首先，先把 f 写成是 n 个概率函数的均值，每个概率函数只有1到2个支撑. 即

$$f(x)=\frac{1}{n}[g_1(x)+g_2(x)+\cdots+g_n(x)], \tag{11.2.8}$$

这里的每个 g_i 为只有1到2个支撑的概率函数. 我们将在例11.2.13中展示怎么去做. 为了模拟 X，首先先模拟一个基于整数 $1,\cdots,n$ 的均匀分布的整数 I(使用例11.2.11的方法)，然后再通过概率函数 g_I 来模拟 X. 读者可以在习题17中证明 X 有概率函数 f.

723

例 11.2.13 使用“Alias”方法模拟二项分布的随机变量 假定要模拟许多来自于参数为9和0.4的二项分布随机变量. 该分布的概率函数 f 在本书中最后的表格中已经给出. 该分布以正概率取 $n=10$ 个不同的数值. 由于 n 个概率之和必须等于1，必定有 x_1 和 y_1 使得 $f(x_1)\leqslant 1/n$ 及 $f(y_1)\geqslant 1/n$. 例如，对 $x_1=0$，$y_1=2$，有 $f(x_1)=0.0101$，而 $f(y_1)=0.1612$. 定义第一个两点概率函数 g_1：

$$g_1(x)=\begin{cases} nf(x_1) & 若\ x=x_1, \\ 1-nf(x_1) & 若\ x=y_1, \\ 0 & 其他. \end{cases}$$

在此例中，$g_1(0)=0.101$ 及 $g_1(2)=0.899$. 然后把 f 写成 $f(x)=g_1(x)/n+f_1^*(x)$，其中：

$$f_1^*(x)=\begin{cases} 0 & x=x_1, \\ f(y_1)-g_1(y_1)/n & x=y_1, \\ f(x) & 其他. \end{cases}$$

在我们的例子中，$f_1^*(2)=0.0173$. 现在，f_1^* 只有 $n-1$ 个值上取正值，而 f_1^* 的正值之和为 $(n-1)/n$. 因此必定存在 x_2 和 y_2 使得 $f_1^*(x_2)\leqslant 1/n$ 及 $f_1^*(y_2)\geqslant 1/n$. 例如，当 $x_2=2, y_2=3$ 时，$f_1^*(x_2)=0.0713$，$f_1^*(y_2)=0.2508$. 把 g_2 定义成：

$$g_2(x)=\begin{cases}nf_1^*(x_2) & 若\ x=x_2,\\ 1-nf_1^*(x_2) & 若\ x=y_2,\\ 0 & 其他.\end{cases}$$

这里，$g_2(2)=0.713$. 可令 $f_1^*(x)=g_2(x)/n+f_2^*(x)$，其中

$$f_2^*(x)=\begin{cases}0 & 若\ x=x_2,\\ f_1^*(y_2)-g_2(y_2)/n & 若\ x=y_2,\\ f_1^*(x) & 其他.\end{cases}$$

在此例中，$f_2^*(3)=0.2221$. 现在 f_2^* 只取 $n-2$ 个正值，总值为 $(n-2)/n$. 再重复该过程 $n-3$ 次，可得到 $g_1,\cdots,g_{n-1}$ 和 f_{n-1}^*. 其中 $f_{n-1}^*(x)$ 只取一个正值，即 $x=x_n$，$f_{n-1}^*(x_n)=1/n$. 使 g_n 成为 x_n 上的一个退化分布．那么对于所有的 x 有 $f(x)=[g_1(x)+\cdots+g_n(x)]/n$.

在这些初步的设定后．"Alias"方法可以快速地进行模拟．模拟独立的 U 和 I，U 服从在区间 $[0,1]$ 上的均匀分布，I 服从整数 $1,\cdots,n$（在我们的例题中 $n=10$）上的均匀分布．如果 $U\leqslant g_I(x_I)$，则令 $X=x_I$；如果 $U>g_I(x_I)$，令 $X=y_I$，这里需要模拟的数据是：

724

i	1	2	3	4	5	6	7	8	9	10
x_i	0	2	1	6	7	3	8	9	4	5
y_i	2	3	3	3	3	4	4	4	5	–
$g_i(x_i)$	0.101	0.713	0.605	0.743	0.212	0.781	0.035	0.003	0.327	1

还有一个更好的方法是用一个简单的模拟来代替 U 和 I 的两个模拟．模拟来自于 $[0,1]$ 上的均匀分布 Y，令 I 小于或等于 $nY+1$ 的最大整数．然后令 $U=nY+1-I$（见习题 19）.

例如，假定模拟服从 $[0,1]$ 上的均匀分布的 Y. 有 $Y=0.4694$. 那么 $I=5$，$U=0.694$. 由于 $0.694>g_5(x_5)=0.212$. 令 $X=y_5=3$. 图 11-4 显示了通过"Alias"方法得到的 1000 组模拟值的直方图. ◀

上面所有构造"Alias"方法所必须的步骤只有在需要模拟许多个来自于同一个离散分布的随机变量时才有价值.

11.2.7 小结

我们已经看到了一些将服从均匀分布的伪随机变量转换为其他分布的伪随机变量的例子．这种接受/拒绝法应用很广，但可能在得到一个被接受的抽样之前需要进行大量的舍去抽样，而且也看到了如何模拟作为其他随机变量函数的随机变量（例如非中心 t 分布随机变量）．另外一些例子告诉我们怎样使用具有一些相同分布的模拟

随机变量.

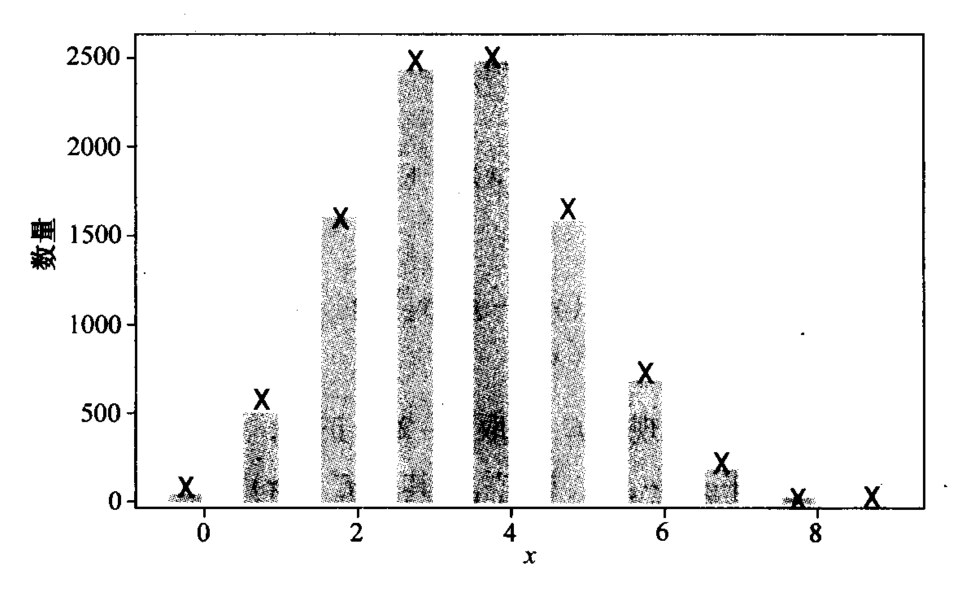


图 11-4　例题 11.2.13 中 10 000 个模拟的二项分布随机变量直方图. ×的记号出现在高度与 10 000 $f(x)$ 相等的地方以此来解释模拟和真实分布一致

725

习题

1. 回到 11.1 节中的习题 10，现在已经知道怎样模拟指数分布随机变量，对习题进行下面的模拟：
   a. 进行 $v_0=2000$ 的模拟，计算 θ 的估计和它的模拟标准误差.
   b. 假设希望 θ 的估计量以 0.99 的概率与 θ 的距离小于 0.01，则需要进行多少次模拟?
2. 请叙述怎样将一组在区间[0,1]上的均匀分布随机样本 $U_1,\cdots,U_n$ 转换为一组容量为 n 的在区间 $[a,b]$ 上的均匀分布的随机样本.
3. 试问怎样使用概率积分变换来模拟服从(11.2.3)式的两个概率密度函数的随机变量.
4. 试问怎样使用概率积分变换来模拟柯西随机变量.
5. 证明：接受/拒绝法中到第一次被接受之前需要迭代的次数的期望为 k.（提示：将每一次迭代看作一个伯努利试验. 在第一次成功之前，试验次数(没有失败)的期望数字是多少?)
6. a. 怎样模拟一个参数为 0 和 1 的 Laplace 分布的随机变量. 参数为 μ 和 σ 的 Laplace 分布的概率密度函数为 $f(x\mid\theta,\sigma)=\dfrac{1}{2\sigma}\mathrm{e}^{-|x-\theta|/\sigma}$.
   b. 怎样先模拟一个 Laplace 随机变量后，再使用接受/拒绝法来模拟一个标准正态随机变量. 提示：求 $x\geqslant 0$ 时，$\mathrm{e}^{-x^2/2}/\mathrm{e}^{-x}$ 的最大值，注意这两个分布都是关于 0 对称的.
7. 假设可以得到尽可能多的你想要的独立同分布的标准正态伪随机数. 试述如何才能从自由度为 4 和 7 的 F 分布中模拟一个伪随机数.
8. 令 X 和 Y 是独立的随机变量，其中 X 服从自由度为 5 的 t 分布，Y 服从自由度为 7 的 t 分布. 我们对 $\mathrm{E}(|X-Y|)$ 感兴趣.
   a. 模拟 1000 对服从上述联合分布的(X_i,Y_i)，并估计 $\mathrm{E}(|X-Y|)$.

b. 运用这1000个模拟数对来估计$|X-Y|$的方差.

c. 根据所估计的方差，需要模拟多少次才能够使得在置信度为0.99时$E(|X-Y|)$的估计量与实际的均值的误差控制在0.01之内?

9. 试述怎样运用接受/拒绝法来模拟具有如下概率密度函数的随机变量.

$$f(x)=\begin{cases}\frac{4}{3}x & \text{若 } 0<x\leqslant 0.5,\\ \frac{2}{3} & \text{若 } 0.5<x\leqslant 1.5,\\ \frac{8}{3}-\frac{4}{3}x & \text{若 } 1.5<x\leqslant 2,\\ 0 & \text{其他.}\end{cases}$$

10. 将例11.1.3的模拟用于例2.1.3的临床试验上. 模拟5000个参数向量. 运用$\alpha_0=1,\beta_0=1$的先验分布. 估计使用“丙咪嗪”组不再复发的概率是最大的概率. 计算需要多少次模拟才能使得在置信度为0.95的情况下，估计量与实际值的概率之间的误差小于0.01.

11. 在例11.2.7，首先模拟了参数为$(n-p)/2$和1的伽玛随机变量τ，假设统计软件允许我们模拟χ^2随机变量. 应该运用哪个χ^2分布，同时如何才能将模拟得到的χ^2分布转为适当的伽玛分布的随机变量.

12. 运用8.4节习题10中表8-1的血压数据. 假设没有把握证明所要处理的两组数据是否有相同的方差. 当允许方差不等时，如在例11.2.8所做的那种，模拟参数的后验分布的样本.

a. 画出两组均值差距的绝对值的样本分布函数的图像.

b. 画出两个方差比值对数的直方图来观察它们的接近程度.

13. 如同(11.2.7)式那样定义F^{-1}. 令U为区间$[0,1]$上的均匀分布. 证明：$F^{-1}(U)$有如同(11.2.6)式的分布函数.

14. 参考在图11-2中的三条曲线，即为三个样本的分布函数$G_{v,1}$，$G_{v,2}$，$G_{v,3}$，称它们所要估计的分布函数分别为G_1，G_2和G_3. 使用Glivenko-Cantelli引理来证明：

$$\Pr(|G_{v,i}(x)-G_i(x)|\leqslant 0.0082,\text{对所有的 } x \text{ 和 } i)$$

[726]

大约为0.9979或更大些. 提示：如果$A_1,\cdots,A_n$是事件，证明：$\Pr(\bigcap_{i=1}^{k}A_i)\geqslant 1-\sum_{i=1}^{k}[1-\Pr(A_i)]$，该不等式就是著名的Bonferroni不等式.

15. 证明：接受/拒绝法可用于离散分布. 即，令f和g为概率函数，而不是概率密度函数，但是接受/拒绝法接下来的部分则不变. 提示：证明可以转化为对x求和以代替积分，而对u的积分则不需转换.

16. 描述怎样用离散的概率变换来模拟一个均值为θ的泊松伪随机变量.

17. 令f是一个概率函数，并且假设(11.2.8)式成立，其中每一个g_i都是另一个概率函数. 假设X是用紧接着(11.2.8)式后所描述的方法模拟得到的. 证明：X有泊松分布函数f.

18. 使用“Alias”方法来模拟一个均值为5的泊松分布. 用书后面的泊松概率表，并且假定16是泊松随机变量的可能取值中的最大值. 假设所有不等于0，…，15的值的概率之和为$k=16$时的概率值.

19. 令 Y 服从区间[0,1]上有均匀分布. 定义 I 为小于或者等于 $nY+1$ 的最大整数, 并且定义 $U=nY+1-I$. 证明: I 和 U 独立, 且 U 在区间[0,1]上有均匀分布.

727

11.3 基于马尔可夫链的蒙特卡罗方法

11.2 节中提到的用来产生特殊分布的伪随机数的方法对于单变量的分布是最有效的. 它们也可用于多变量场合, 但应用并不广泛. 一种基于马尔可夫链的方法在 Metropolis(1953)和 Gelfand, Smith(1990)的论文发表后变得普及化. 本节我们仅考虑 MCMC 方法的最简单的形式.

11.3.1 吉布斯抽样算法

我们将试着先从模拟一个二元分布开始, 设(X_1,X_2)的联合概率密度函数为 $f(x_1,x_2)=cg(x_1,x_2)$. 假定函数 g 已知, 而未必知道常数 c, 该种情况常在计算后验分布时遇到. 如果把 X_1, X_2 看作是参数, 函数 g 可能是先验概率密度函数与似然函数的乘积 (其中 X_1, X_2 的取值看成已知的). 由于 $cg(x_1,x_2)$为后验分布的概率密度函数, 因此常数 $c=1/\int g(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2$, 但是 c 一般很难算出. 即使我们能大致地确定 c, 但后验分布中还是有一些特性不是很简单就能算出的, 所以模拟的方法是有用的.

如果函数 $g(x_1,x_2)$有特殊的形式, 则可用一种有效的算法模拟具有概率密度函数 f 的随机向量. 下面具体描述所要求的这种形式: 首先固定 x_2, 把 $g(x_1,x_2)$看作 x_1 的函数, 该函数必须看起来是一个可以模拟出伪随机数的概率密度函数 (对 X_1 而言). 类似地, 如果固定 x_1, 把 $g(x_1,x_2)$看作 x_2 的函数, 该函数也必须看起来是一个可以模拟出伪随机数 X_2 的概率密度函数.

例 11.3.1 从正态分布中抽样 设有一个来自于均值 μ 和标准差 τ 未知的正态分布的样本. 其先验和似然函数乘积由下式给出, 为了方便起见没有给出其中的常数因子, 该表达式为:

735

$$\xi(\mu,\tau|x)\propto\tau^{\alpha_1+1/2-1}\exp\left(-\tau\left[\frac{1}{2}\lambda_1(\mu-\mu_1)^2+\beta_1\right]\right),$$

其中当观察数据已知时, α_1, β_1, μ_1, λ_1 均已知. 对于固定的 τ, 将上式看作是 μ 的函数, 该函数看起来是服从均值为 μ_1 且方差为$(\tau\lambda_1)^{-1}$的正态分布的概率密度函数. 对于固定的 μ, 把它看作是 τ 的函数, 它看起来像是参数为 $\alpha_1+1/2$ 和 $\lambda_1(\mu-\mu_1)^2/2+\beta_1$ 的伽玛分布的概率密度函数, 而这两个分布都比较容易模拟. ——◀

当对于固定的 x_2, 将 $g(x_1,x_2)$看作是 x_1 的函数时, 考虑给定 $X_2=x_2$ 的条件下 X_1 的条件密度函数, 而不考虑不依赖于 x_1 的一个常数因子(见习题 1). 类似地, 当对

固定的 x_2，将 $g(x_1,x_2)$看作是 x_1 的函数时，相当于考虑给定 $X_1=x_1$ 时 X_2 的条件密度函数.

一旦确定函数 $g(x_1,x_2)$具有我们想要的形式，那么就有如下的算法步骤：

1. 给定 X_2 的初始值 $x_2^{(0)}$，令 $i=0$.
2. 从给定 $X_2=x_2^{(i)}$ 时 X_1 的条件分布中模拟出一个新值 $x_1^{(i+1)}$.
3. 从给定 $X_1=x_1^{(i+1)}$ 时 X_2 的条件分布中模拟出新值 $x_2^{(i+1)}$.
4. 将 i 用 $i+1$ 替代并回到第 2 步.

该算法循环至一个足够大的 i 值时结束. 虽然现在还没有一个真正令人满意的收敛准则，在本节的后面将介绍一个收敛准则，通常称该算法为吉布斯抽样. 这个名字来自于早期进行抽样的 Geman and Geman(1984)，他们抽样时所用的分布为吉布斯分布.

11.3.2 一些理论准则

迄今为止，还没有一个判断吉布斯抽样算法好坏的准则. 该准则来源于这样的事实，即一列$(x_1^{(1)},x_2^{(1)}),(x_1^{(2)},x_2^{(2)}),\cdots$构成了马尔可夫链的状态序列. 该马尔可夫链比在 2.4 节中遇到的更加复杂，原因有二：第一，状态是二维的；第二，状态的个数是无限的而不是有限的. 尽管如此，还是容易确认描述吉布斯抽样算法的马尔可夫链的基本结构. 假设 i 是迭代步数的现值，在给定已有的状态$(X_1^{(1)},X_2^{(1)}),\cdots,(X_1^{(i)},X_2^{(i)})$的条件下，下一个状态$(X_1^{(i+1)},X_2^{(i+1)})$的条件分布仅依赖于当前状态$(X_1^{(i)},X_2^{(i)})$. 因而与 2.4 节所给的有限马尔可夫链定义的性质几乎一样. 这里在给定过去和现在的状态下，接下来的状态的条件分布必须仅依赖于当前状态. 吉布斯抽样算法则将该特征推广到可能具有无限多的二元状态的情形.

即使我们同意该抽样序列成立了，那么凭什么相信它们就是来自于所想要的分布呢？答案来自于根据马尔可夫链多次迭代的一些困难的数学结果. 可以用一个简单例子来说明.

736

例 11.3.2 马尔可夫链转移概率的收敛性 考虑一个具有三个状态的有限马尔可夫链，其转移矩阵为

$$\boldsymbol{P}=\begin{pmatrix}0.2 & 0.7 & 0.1\\ 0.8 & 0.1 & 0.1\\ 0.3 & 0.3 & 0.4\end{pmatrix}.$$

在 2.4 节中知道了该马尔可夫链的 m 步转移矩阵为 $\boldsymbol{P}^m$. 该矩阵给出了马尔可夫链的两个状态的经 m 步的转移矩阵. 例如，当 $m=20$，由上面的矩阵 $\boldsymbol{P}$，我们有

$$\boldsymbol{P}^{20}=\begin{pmatrix}0.4553 & 0.4018 & 0.1429\\ 0.4553 & 0.4018 & 0.1429\\ 0.4553 & 0.4018 & 0.1429\end{pmatrix}.$$

注意到现在P^{20}的三行要比P的三行彼此接近程度更高. 如果我们的初始概率向量w是任意的，那么必有wP^{20}为$w^* = (0.4553, 0.4018, 0.1429)$. 即不管马尔可夫链初始状态有什么分布，经过20步的迭代后它的分布都将为w^*，且在本例中很容易证明向量w^*有一特殊性质：$w^*P = w^*$. ◀

如果矩阵P的每一元素均大于0，那么例11.4.2中的结论对于所有的有限马尔可夫链都成立(事实上,比这个更一般). 通常只要P^m收敛(当$m \to \infty$)到一个矩阵，该矩阵的各行都是同一个行向量，记为w^*，且该向量满足$w^*P = w^*$，那么本结论成立. 在我们的例子中，P^{20}与极限的差距小于小数点后四位. 向量w^*被称作为平稳分布，因为它有性质：如果w^*是状态在i时的分布，那么w^*也同样是以后所有状态的分布. 甚至于即使w^*不是状态在i时的分布，但是由于收敛性，只要m足够大，$i+m$时状态的分布必将与w^*非常接近. 事实上，当应用于更一般的马尔可夫链时，正是吉布斯抽样算法所做的那样. 无论马尔可夫链的初始分布如何，在足够大的迭代步数之后，该分布都将近似于一个平稳分布. 而且也容易证明想要模拟的分布正是该平稳分布. 举个例子，如果$(X_1^{(i)}, X_2^{(i)})$本身有自己的联合概率密度函数f，那么$(X_1^{(i+1)}, X_2^{(i+1)})$也会有这样的概率密度函数，而且$(X_1^{(i+1)}, X_2^{(i+1)})$(见习题2)也一样具有此密度函数. 由于吉布斯抽样与马尔可夫链的紧密联系，因此通常称吉布斯抽样方法和其他一些相关的方法为MCMC.

注： 马尔可夫链的收敛需要一定的条件. 不是所有的马尔可夫链都有平稳分布，也不是所有的有平稳分布的马尔可夫链都收敛. 然而，在这里能描述的所有简单例子中，收敛性都是成立的. 更一般的讨论可参见Tierney(1994).

11.3.3　马尔可夫链何时收敛?

737

为了从马尔可夫链的平稳分布中获得一个样本，需要等到马尔可夫链收敛的时候，当然在理论上讲，这需要永远等下去. 然而，正如在例11.3.2中所见，可以在有限的时间内得到一个比较好的近似. 但是在特定的应用时，怎么来解释马尔可夫链是否足够长，使得我们确信抽得的样本可以看作是从平稳分布中得到的呢? 有很多工作围绕这个问题展开. 但并没有一个简单的证明方法. Cowles和Carlin(1996)回顾了一些在蒙特卡罗分析中评价马尔可夫链的收敛性的方法，这里给出一个简单的技术.

考虑k个不同的马尔可夫链，它们的初始值分别为$x_{2,1}^{(0)}, \cdots, x_{2,k}^{(0)}$. 这$k$个马尔可夫链不仅用于评估收敛性，而且用于计算模拟估计的方差. 比较明智的选择是将初始值$x_{2,1}^{(0)}, \cdots, x_{2,k}^{(0)}$尽可能分散开. 这将有助于判断是否得到一个收敛得很慢的马尔可夫链. 接下来应用吉布斯抽样算法，从k个初始值的每一个开始. 这样我们将有k个独立的马尔可夫链，每个都有相同的平稳分布. 如果k个马尔可夫链经过m次迭代来抽样，那么我们就可以把X_1(或X_2)的观察值看作为k个大小均为m的样本. 为了记号的方便，令$T_{i,j}$代表X_1或X_2的来自于第i个马尔可夫链的第j次迭代的值. (我们

将对 X_1 和 X_2 交错地重复以下的分析.)现在，将 $T_{i,j}$，$j=1,\cdots,m$ 看作为来自于 k 个分布中第 i 个($i=1,\cdots,k$)分布的容量为 m 的一个样本. 如果我们进行足够长的抽样以使得马尔可夫链近似收敛，那么所有 k 个分布应该是近似相同的. 这意味着可以用方差分析(10.3 节)中的 F 统计量来度量这 k 个分布的接近程度. 这个 F 统计量可以写成 $F=B/W$，其中

$$B=\frac{m}{k-1}\sum_{i=1}^{k}(\bar{T}_{i+}-\bar{T}_{++})^2,$$

$$W=\frac{1}{k(m-1)}\sum_{i=1}^{k}\sum_{j=1}^{m}(T_{ij}-\bar{T}_{i+})^2.$$

这里我们使用了与 10.3 节相同的记号，下标 + 出现的位置表示对该位置上所有下标对应的值取平均. 如果这 k 个分布是不同的，那么 F 值会很大. 如果 k 个分布都相同，那么 F 值应该接近于1. 如前所述，我们计算了两个 F 统计量，一个基于 X_1，另一个基于 X_2. 如果两个 F 统计量都小于某个略大于 1 的数，则认为已经进行了足够长度的抽样. Gelman 等(1995)从本质上描述了这个同样的过程，并建议将两个 F 统计量中大的那个与 $1+0.44m$ 比较. 在计算 F 统计量之前，至少从 $m=100$ 开始（如果迭代速度足够快的话）是一个比较好的主意. 这将有助于避免仅由于某些“幸运”的模拟结果而宣称已经收敛了. 在确定收敛之前，马尔可夫链迭代的初始序列常被称作为“曝光”. 在曝光迭代之后，接下来的迭代则可认为来自于平稳分布. 然而，由于一个马尔可夫链的迭代是相关的，因此不应该把它们看作为独立同分布的样本. 由于不能确保曝光迭代的分布是否接近于平稳分布，因此通常将其舍去. 738

例 11.3.3 新墨西哥的护理院 我们使用来自 1998 年新墨西哥 18 个非乡村护理院中护理病人的天数的数据. 这里认为观察值为来自均值 μ 和方差 τ 均未知的正态分布的随机样本. 设使用自然共轭先验，已计算出后验超参数为 $\alpha_1=11$, $\beta_1=50\,925.37$, $\mu_1=183.95$ 及 $\lambda_1=20$. 我们以例 11.3.1 中的吉布斯抽样算法为例说明上述收敛判断. 正如在例 11.3.1 中发现的一样，给定 τ 下 μ 的条件分布为均值是 183.95、方差是 $(20\tau)^{-1}$ 的正态分布，给定 μ 下 τ 的条件分布是参数为 11.5 和 $50\,925.37+20(\mu-183.95)^2$ 的伽玛分布. 对于 μ 选择 $k=5$ 个初始值：182.17, 227, 272, 137, 82. 选择这些初始值时，先得到 μ 的后验标准差的一个近似：$(\beta_1/[\lambda_1\alpha_1])^{1/2}\approx 15$. 然后取后验均值和与后验均值上下相差 3 个和 6 个上述的后验标准差的值构成. 在能计算 F 统计量之前必须对这 5 个马尔可夫链进行了 $m=2$ 次迭代. 在我们的模拟中，当 $m=2$ 时，两个 F 统计量中较大的值已经小到 0.8862 了，在到 $m=100$ 的过程中，一直非常接近于 1 了，此时，很显然应该停止曝光阶段了. ◀

11.3.4 基于吉布斯抽样的估计

迄今为止，我们已经讨论了(没有证明)如果运行吉布斯抽样算法，进行多次的迭代(经过曝光)，应该开始看到其联合概率密度接近于想要抽样的 f 函数的($X_1^{(i)}$,

$X_2^{(i)}$)对了. 不幸的是，这些相继的数对即使有相同的分布，但也不是相互独立的. 大数定律并没有告诉我们相关随机变量的平均值会收敛. 然而，由马尔可夫链中得到的这种相关类型是充分规则的，以至于有一些定理保证了平均值的收敛性，甚至这些平均值是渐近正态的. 即，假如希望基于马尔可夫链的 m 个观察值来估计某个函数 $h(X_1,X_2)$ 的均值 μ，仍可以假定$\frac{1}{m}\sum_{i=1}^{m} h(X_1^{(i)},X_2^{(i)})$收敛于 μ，且有均值为 μ 方差为 σ^2/m 的渐近正态分布. 但是，此时的收敛速度明显慢于独立同分布样本的收敛速度，而且 σ^2 也比 $h(X_1,X_2)$ 的方差大，这是由于 $h(X_1^{(i)},X_2^{(i)})$ 的序列值通常是正相关的. 正相关同分布的随机变量的平均值的方差比由相同数量的独立同分布的随机变量的平均值的方差要大（习题4).

我们将使用用于确定曝光迭代的相同的 k 个独立的马尔可夫链来处理由于相关样本带来的问题. 去掉曝光部分，继续对每个马尔可夫链中各自再进行 m_0 次迭代，对每个马尔可夫链计算想要的估计，可能是均值、样本分位数、样本方差或其他一些测度，$Z_j, j=1,\cdots,k$，然后计算式(11.1.2)中的 S，

739

$$S=\left(\frac{1}{k}\sum_{j=1}^{k}(Z_j-\bar{Z})^2\right)^{1/2}. \tag{11.3.1}$$

得到 Z_j 的模拟方差的估计值 S^2. 记模拟方差为 σ^2/m_0，并像例11.1.9那样，用 $\hat{\sigma}^2=m_0S^2$ 来估计 σ^2. 同时，将由 k 个链中得到的所有样本组合成单一的样本，然后用这单一样本构造总的估计 Z. 估计 Z 的模拟标准差为$(\hat{\sigma}/(m_0k))^{1/2}=S/k^{1/2}$.

此外，我们可能希望确定获得一个精确估计所需的模拟次数. 用 $\hat{\sigma}$ 来替换(11.1.5)中的 σ 来得到一个适当的模拟次数 v. 将这 v 次模拟分配到 k 个马尔可夫链中，当 $v/k>m_0$ 时，每条链至少将迭代 v/k 次.

11.3.5 一些例子

例 11.3.4 新墨西哥的护理院 我们不需要用吉布斯抽样来模拟服从例11.3.1中的后验分布的样本，因为在本例中的 μ 和 τ 的联合分布有一个显式表达式. 每个边际分布和条件分布都已知且易于模拟. 只有当条件已知时，吉布斯抽样是最有用的. 然而，可以说明例11.3.1中吉布斯抽样的使用，并将模拟结果与 μ 和 τ 的已知边际分布相比较.

在例11.3.3中，使用了 $k=5$ 个马尔可夫链并对它进行100次迭代曝光. 现在希望可以从联合后验分布中构造出(μ,τ)数对的一个样本. 曝光后，对每一个马尔可夫链再进行 $m_0=1000$ 次迭代. 这些迭代将产生数对(μ,τ)的5个相关序列. 对 μ 值的序列对之间的相关性很小. 对于 τ 的数对序列也一样. 为了将结果与已知的后验分布相比较，图11-5显示了 μ 的 t 分位数点和 τ 的伽玛分位数点图(在前面已经介绍过正态分位数点图，也可以用相同的方法来构造伽玛和t分位数点图，只

要用伽玛和 τ 分位数代替标准分位数即可)．在图 11-5 中模拟值看起来与所画的直线很接近（尾部有些点偏离了直线，但事实上所有分位点图都会发生这种情况）．图 11-5 显示了由真实后验分布得到的分位数线，其中对 μ 而言是由自由度为 22 的 t 分布得到，并乘以 15.21，其中心在 183.95 处，对 τ 而言是由参数为 11 和 50 925.7 的伽玛分布得到的．

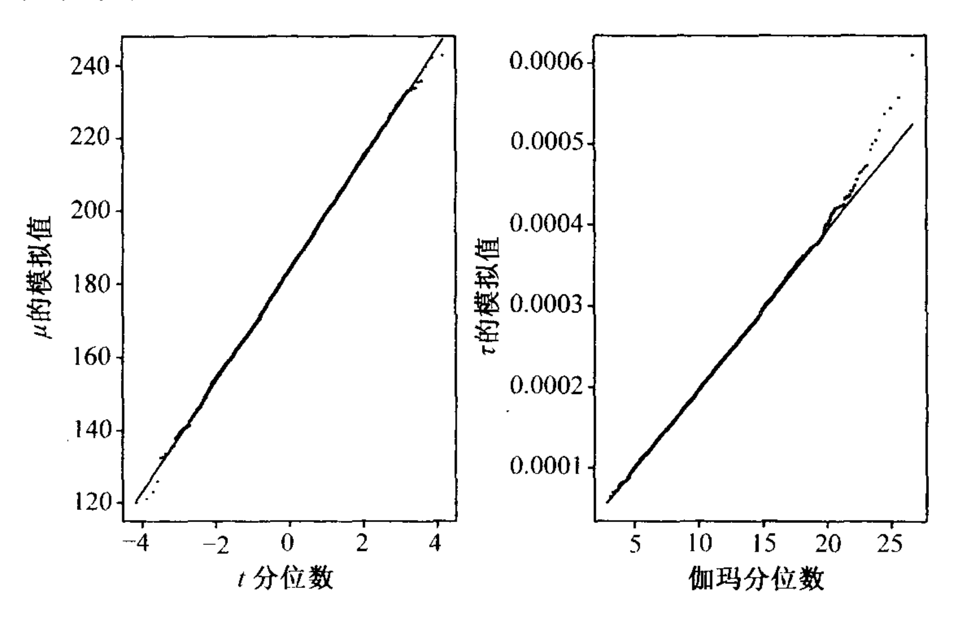


图 11-5 从例 11.3.4 中后验分布模拟得到的 μ 和 τ 的分位数图．每个图中的直线表示得到的真实后验分布的分位数．左图中的横轴为自由度为 22 的 t 分布的分位数，μ 的真实后验分布做了尺度和位移变换后变为此 t 分布．右图中的横轴是参数为 11 和 1 的伽玛分布的分位数．τ 的真实后验分布是该伽玛分布做一个尺度变换

可以用 (μ,τ) 对的样本来估计 (μ,τ) 的任何函数的后验均值．例如，假定我们对 $\mu+1.645/\tau^{1/2}$ 的均值 θ 感兴趣，它是未知分布的原始观察值的 0.95 分位数，对 $\mu+1.645/\tau^{1/2}$ 的 5000 个模拟值的平均值为 $Z=299.67$．(11.3.1) 式中的 S 值为 0.4119，可得 $\sigma=13.03$．从而 Z 的模拟标准误差为 $\hat{\sigma}/5000^{1/2}=0.1842$．在本例中，可以准确地计算 $\mu+1.645/\tau^{1/2}$ 的真正后验均值，这就是

$$\mu_1+1.645\beta_1^{1/2}\frac{\Gamma(\alpha_1-0.5)}{\Gamma(\alpha_1)}=299.88.$$

它与 Z 的模拟值偏离稍大于 1 个拟合标准差．假定希望 θ 的估计值与真实值的差距小于 0.01 的概率达到 0.99．将这些值以及 $\hat{\sigma}=13.03$ 代入 (11.1.5) 式中，发现总共需要进行 $v=12\ 358\ 425$ 次模拟．5 个马尔可夫链中的每一个都必须进行 2 251 685 次的迭代． ◀

740

当遇到两个以上的参数问题时，吉布斯抽样才真正体现了它的价值．概率密度函数为 $f(\boldsymbol{x})=cg(\boldsymbol{x})$ 的 p 元随机变量 $(X_1,\cdots,X_p)$ 的一般吉布斯抽样算法如下：首先，必须保证当其他所有变量给定时，g 作为单个变量的概率密度函数似乎是容易被模拟

的. 然后进行下面的步骤:

1. 为 $X_2, \cdots, X_p$ 选初始值 $x_2^{(0)}, \cdots, x_p^{(0)}$, 并令 $i=0$.

2. 从给定 $X_2 = X_2^{(i)}, \cdots, X_p = x_p^{(i)}$ 时 X_1 的条件分布中模拟出新值 $x_1^{(i+1)}$.

3. 从给定 $X_1 = x_1^{(i+1)}, X_3 = x_3^{(i)}, \cdots, X_p = x_p^{(i)}$ 时 X_2 的条件分布中模拟出新值 $x_2^{(i+1)}$.

⋮

$p+1$. 从给定 $X_1 = x_1^{(i+1)}, \cdots, X_{p-1} = x_{p-1}^{(i+1)}$ 时 X_p 的条件分布中模拟出新值 $x_p^{(i+1)}$.

741 $p+2$. 用 $i+1$ 代替 i, 再回到步骤 2.

根据该算法得到的 p 元 $(X_1, \cdots, X_p)$ 值的序列是与以前意义相同的马尔可夫链. 该马尔可夫链的平稳分布有概率密度函数 f, 开始后经过多步迭代, 每个迭代的分布接近于该平稳分布.

例 11.3.5 有不合适的先验分布的多元回归 考虑一个三元组 (Y_i, x_{1i}, x_{2i}), $i=1, \cdots, n$ 的观察数据的问题. 假设 x_{ji} 的值是已知的, Y_i 的分布服从均值为 $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$、精度为 τ 的正态分布. 这是多元回归模型, 其中方差用精度的倒数代替. 假定对参数采用了不正确的先验分布 $\xi(\beta_0, \beta_1, \beta_2, \tau) = 1/\tau$. 那么参数的后验概率密度函数与似然函数的 $1/\tau$ 倍数成比例, 即为某个常数乘以

$$\tau^{n/2-1} \exp\left(-\frac{\tau}{2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 \right). \tag{11.3.2}$$

为简化已有的公式, 先定义数据的一些描述性统计量.

$$\bar{x}_1 = \frac{1}{n} \sum_{i=1}^{n} x_{1i}, \bar{x}_2 = \frac{1}{n} \sum_{i=1}^{n} x_{2i}, \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

$$s_{11} = \sum_{i=1}^{n} x_{1i}^2, s_{22} = \sum_{i=1}^{n} x_{2i}^2, s_{12} = \sum_{i=1}^{n} x_{1i} x_{2i},$$

$$s_{1y} = \sum_{i=1}^{n} x_{1i} y_i, s_{2y} = \sum_{i=1}^{n} x_{2i} y_i, s_{yy} = \sum_{i=1}^{n} y_i^2.$$

对于给定的 β_0, β_1, β_2 的值, 将(11.3.2)式看作是 τ 的函数, 它看起来是参数为 $n/2$ 和 $\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 / 2$ 的伽玛分布的概率密度函数. 将(11.3.2)式看成给定其他参数时 β_j 的函数, 它是 β_j 的指数函数, 指数为 β_j 的二次方程, 且 β_j^2 项的系数为负. 如此, 它看起来像是一个均值依赖于数据和其他 β、方差等于 $1/\tau$ 乘以数据的某个函数的正态随机变量的概率密度函数. 通过对表达式 $\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2$ 中的平方分别做三次计算, 每次将不同的 β_j 视为感兴趣的变量, 该表达式将变得更清晰. 例如, 将 β_0 视为感兴趣的变量, 可得到

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 = n(\beta_0 - [\bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2])^2,$$

加上一个与 β_0 无关的项. 所以，给定其他参数下 β_0 的条件分布是一个均值为 $\bar{y}-\beta_1\bar{x}_1-\beta_2\bar{x}_2$、方差为 $1/[n\tau]$ 的正态分布. 如将 β_1 视为感兴趣变量，可得到

$$\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{1i}-\beta_2x_{2i})^2=s_{11}(\beta_1-w_1)^2,$$

742

加上一个与 β_1 无关的项，其中

$$w_1=\frac{1}{s_{11}}(s_{1y}-\beta_0n\bar{x}_1-\beta_2s_{12}).$$

这表明给定其他参数的情况下 β_1 的条件分布是一个均值为 w_1、方差为 $(\tau s_{11})^{-1}$ 的正态分布. 类似地，给定其他参数的情况下 β_2 的条件分布是一个均值为 w_2、方差为 $(\tau s_{22})^{-1}$ 的正态分布，其中

$$w_2=\frac{1}{s_{22}}(s_{2y}-\beta_0n\bar{x}_2-\beta_1s_{12}).$$ ◀

例 11.3.6 20 世纪 50 年代的失业率 如表 11-4 所示，我们看到 1951 ~ 1959 年的失业率数据比包含 1950 年失业率的数据看起来更满足多元回归的假设.

表 11-4 例 11.3.6 的失业数据

失业	生产指数	年限
3.1	113	1950
1.9	123	1951
1.7	127	1952
1.6	138	1953
3.2	130	1954
2.7	146	1955
2.6	151	1956
2.9	152	1957
4.7	141	1958
3.8	159	1959

我们将用一个不恰当的先验分布和吉布斯抽样来得到这些参数的后验分布的样本. 例 11.3.5 已给出所需的条件分布，只需描述性统计量的值，并取 $n=9$：

$$\bar{x}_1=140.7778,\quad \bar{x}_2=6,\quad \bar{y}=2.789,$$
$$s_{11}=179\,585,\quad s_{22}=384,\quad s_{12}=7837,$$
$$s_{1y}=3580.9,\quad s_{2y}=169.2,\quad s_{yy}=78.29.$$

再次运行 $k=5$ 个马尔可夫链. 在此问题中，有四个参数坐标：β_i，$i=0$，1，2 及 τ. 这样要计算四个 F 统计量，并在最大的 F 小于 $1+0.44m$ 前进行曝光处理. 假定这会在 $m=4546$ 时发生，然后对每个马尔可夫链再进行 10 000 次的迭代抽样.

假定我们想要包含 β_1 后验分布 90% 的一个区间 $[a,b]$. 数 a，b 分别是样本的 0.05 和 0.95 的分位数. 基于得到的 β_1 的总共 5000 个样本，此区间为 $[-0.1178,-0.0553]$.

为了评估区间端点的不确定性，对五个马尔可夫链中的每一个都计算 0.05 和 0.95 的样本分位数. 那些值分别为：

0.05 分位数：-0.1452，-0.1067，-0.1181，-0.1079，-0.1142

0.95 分位数：-0.0684，-0.0610，-0.0486，-0.0594，-0.0430

基于样本 0.05 分位数，S 的值为 0.015 67，且基于样本 0.95 分位数，S 的值为 0.011 42. 为安全起见，用这两个值中较大的那个值来估计区间端点的模拟标准误差，由于每个链都运行了 $m_0 = 10\,000$ 次迭代，于是就有 $\hat{\sigma} = Sm_0^{1/2} = 1.567$. 如果希望以 0.95 的概率使得区间的每个端点与 β_1 的分布真实分位数的差距在 0.01 范围内（每个端点都在 0.01 范围内的概率应该更小些，但计算比较困难）. 可以用 (11.1.5) 式来计算需要多少模拟值，得到 $\nu = 94\,386$，这就意味着 5 个链的每一个都需要进行 18 878 次的迭代，这大约是已经模拟次数的 2 倍. 作为比较，构建的 β_1 的 90% 的置信区间是 [-0.1124, -0.0579]. 这与后验概率区间非常地接近. ◀

743

可以找到例 11.3.6 的后验分布的显式形式，虽然在本书中并没有这样做. 事实上，当使用不合适的先验时，在本例的最后计算的 90% 的置信区间包含了 90% 的后验分布概率，与参数 $1-\alpha_0$ 的置信区间包含了 $1-\alpha_0$ 的后验概率是一样的. 下面是个显式解不存在的例子.

例 11.3.7　有不同方差的一维空间的贝叶斯分析　考虑一维空间，其中假设数据中每个观测值都来自于具有不同均值但有相同方差的 p 维正态分布，为了说明吉布斯抽样进一步的功能，取消每个正态分布有相同方差的假设，即，对于 $i = 1, \cdots, p$，假设 $Y_{i1}, \cdots, Y_{in_i}$ 都服从于均值为 μ_i 和精度为 τ_i 的正态分布. 且所有观测值对于所有参数都是条件独立的. 参数的先验分布如下：假设给定其他参数 $\mu_1, \cdots, \mu_p$ 条件下独立，且 μ_i 服从均值为 ψ 和精度为 $\lambda_0\tau_i$ 的正态分布. 其中，ψ 是另一个服从某种分布的参数. 引入参数 ψ 是为了说明 μ_i 都是来自于一个公共的分布，但并不想确定是何种分布. 然后说 ψ 是服从均值为 ψ_0、精度为 u_0 的正态分布，对于不恰当的先验分布，下面令 $u_0 = 0$，因而就不需要 ψ_0 了. 然后设 $\tau_1, \cdots, \tau_p$ 是服从参数为 α_0 和 β_0 的伽玛分布. ψ 与 τ_i 相互独立. 对于不恰当的先验，令 $\alpha_0 = \beta_0 = 0$. 上述的这个模型称之为层次模型，因为分布有一个水平的层次. 图 11-6 说明了本例中的层次.

观察值和参数的联合概率密度函数是似然函数的乘积（给定 μ_i 和 τ_i 下观察值的概率密度函数）乘以给定 τ_i 和 ψ 下 μ_i 的条件先验概率密度函数，再乘以 τ_i 和 ψ 的先验概率密度函数. 除了与数据和参数无关的常数外，该乘积有如下形式：

$$\exp\left(-\frac{u_0(\psi-\psi_0)^2}{2} - \sum_{i=1}^{p}\tau_i\left[\beta_0 + \frac{n_i(\mu_i-\bar{y}_i)^2 + w_i + \lambda_0(\mu_i-\psi)^2}{2}\right]\right) \tag{11.3.3}$$

$$\times \prod_{i=1}^{p}\tau_i^{\alpha_0+[n_i+1]/2-1},$$

其中 $w_i = \sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2, i=1,\cdots,p$. 我们已经对于(11.3.3)式进行排列，以便包含每个参数的项都互相接近. 这将有利于描述吉布斯抽样算法.

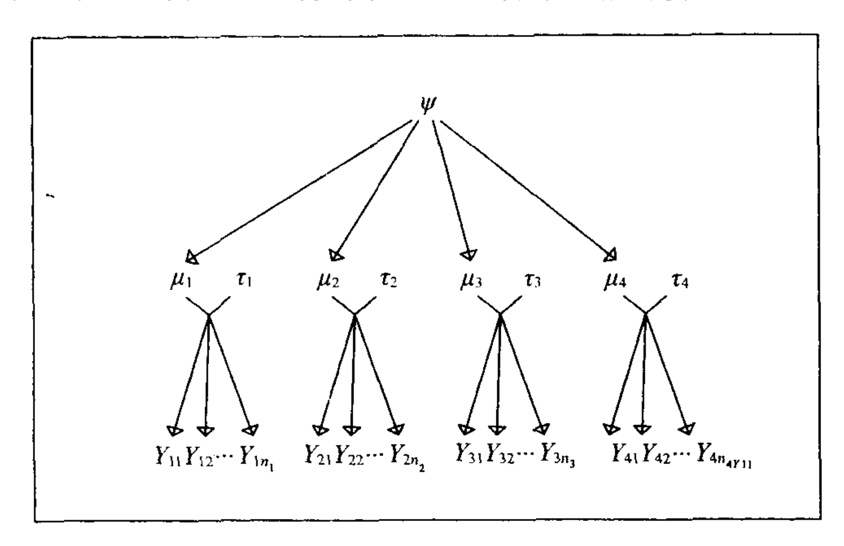


图 11-6 例 11.3.7 中层次模型的图. 参数 ψ 影响 μ_i 的分布，而参数 (μ_i, τ_i) 影响 Y_{ij}的分布

为了构造吉布斯抽样，要把 (11.3.3)式作为每个参数的函数分别加以考察. 参数为 $\mu_1,\cdots,\mu_p$; $\tau_1,\cdots,\tau_p$; 以及 ψ. 作为 τ_i 的函数，(11.3.3)式看起来像是参数为 $\alpha_0+(n_i+1)/2$ 和 $\beta_0+[n_i(\mu_i-\bar{y}_i)^2+w_i+\lambda_0(\mu_i-\psi)^2]/2$ 的伽玛分布的概率密度函数. 作为 ψ 的函数，它像均值为 $[u_0\psi_0+\lambda_0\sum_{i=1}^{p}\tau_i\mu_i]/[u_0+\lambda_0\sum_{i=1}^{p}\tau_i]$ 且精度为 $u_0+\lambda_0\sum_{i=1}^{p}\tau_i$ 的正态分布的概率密度函数. 这是对包含 ψ 的所有项做平方得到的. 类似地，对包含 μ_i 的所有项计算平方后发现(11.3.3)看似均值为 $[n_i\bar{y}_i+\lambda_0\psi]/[n_i+\lambda_0]$和精度为 $\tau_i(n_i+\lambda_0)$ 的正态概率密度函数. 所有这些分布都易于模拟.

744

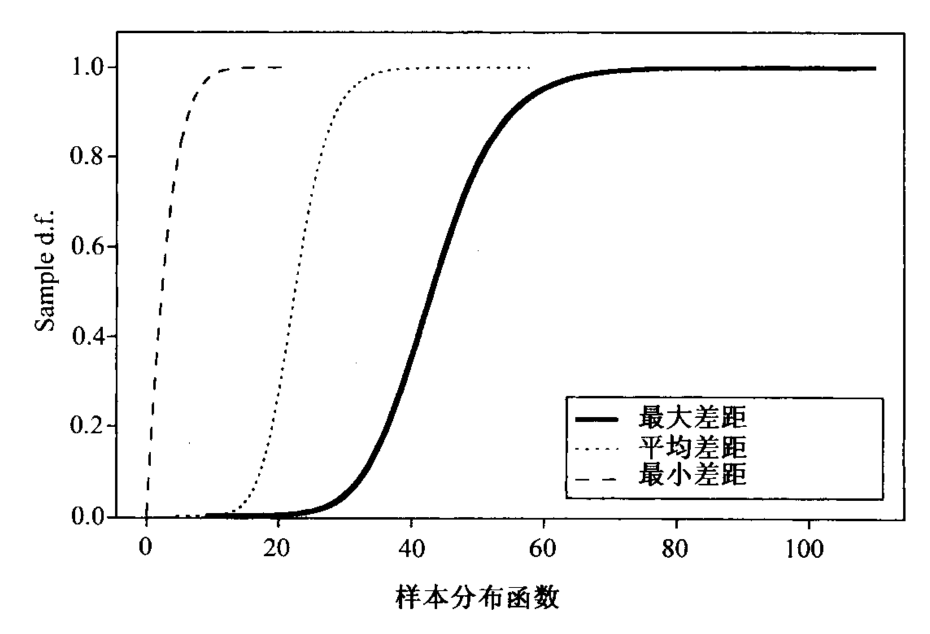
用例 10.3.1 中热狗中卡路里的数据作为例子. 在本例中，$p=4$，我们要用的先验分布中 $\lambda_0=\alpha_0=1, \beta_0=0.1, u_0=0.001$ 和 $\psi_0=170$. 用 $k=6$ 的马尔可夫链和做 $m=100$ 的曝光模拟，为使所有 9 个 F 统计量中的最大值小于 $1+0.44m$. 要运行曝光模拟大大多于所需要的次数，然后对 6 个马尔可夫链再做 10 000 次迭代，来自后验分布的样本足以让我们回答可能有的关于参数的任何问题，包括一些在第 10 章中运用分析方法不能回答的问题. 例如，在表 11-5 中列出的一些参数的后验均值和标准偏差，注意到四组中方差 $1/\tau_i$ 有多大的不同. 例如，通过计算满足 $1/\tau_4^{(l)}>4/\tau_1^{(l)}$ 的迭代数 l 的数量与 60 000 的商，就可以计算 $1/\tau_4>4/\tau_1$ 的概率，其结果是 0.5087. 该数字表明至少某些方差相差很大的可能性是比较大的. 如果方差是不同的，那么第 10 章中的 ANOVA(方差分析)计算就不成立.

表 11-5 例 11.3.7 中的一些要素的后验均值和标准差

种类	牛肉	猪肉	禽肉	特色类
i	1	2	3	4
$E(\mu_i \mid y)$	156.7	158.4	120.7	159.7
$(\mathrm{Var}(\mu_i \mid y))^{1/2}$	3.498	5.241	6.160	10.55
$E(1/\tau_i \mid y)$	252.3	487.3	670.8	1100
$(\mathrm{Var}(1/\tau_i \mid y))^{1/2}$	84.70	179.1	250.6	586.9
$E(\psi \mid y)=152.8$	$(\mathrm{Var}(\psi \mid y))^{1/2}=10.42$			

我们也可能提出这样的问题：μ_i 之间有多大差别. 作为比较，可以做例 11.2.7 中相同的计算. 在 60 000 个模拟的 99% 中至少有一个 $|\mu_i^{(l)}-\mu_j^{(l)}|>24.66$. 大约一半的模拟有 $|\mu_i^{(l)}-\mu_j^{(l)}|>2.268$. 在 99% 的模拟中，差别的平均值至少是 13.07. 图 11-7 给出了六个 $|\mu_i-\mu_j|$ 差值的最大值、最小值和平均值的样本分布函数的图形. 仔细检查本例的结果，显示出：四个 μ_i 比分析例 11.2.7 后所认为的更加靠近. 当我们在分层模型中使用适当的先验分布，这是会发生的典型的事情. 在例 11.2.7 中，μ_i 都是独立的，它们在先验分布中没有一个公共的未知均值. 在例 11.3.7 中，μ_i 都有均值是未知参数 ψ 的公共的先验分布. 该附加参数 ψ 的估计允许 μ_i 的后验分布推向位于所有样本的平均值的附近的位置. 根据该数据，总的样本均值为 147.60. ◀

745



746

图 11-7 例 11.3.7 中六个 $|\mu_i-\mu_j|$ 差别的最大值，平均值和最小值的样本分布函数

11.3.6 预测

本节中所有例子中所做的计算都涉及参数的函数. 从吉布斯样本中获得的后验分布的样本也可以用于对未来观测值做预测和预测区间. 用来预测的最直接的办法

是使用后验样本得到的参数值来拟合未来数据. 虽然对于预测有很多有效的方法，但该方法易于描述和评价.

例 11.3.8 热狗中的卡路里 在例 11.3.7 中，我们也许关心两个热狗的卡路里数有多少差别. 例如，令 Y_1 和 Y_3 分别为牛肉热狗和家禽热狗的卡路里数，可以如下构造 $D = Y_1 - Y_3$ 的预测区间. 对于每一个迭代 l，令模拟的参数向量为

$$\theta^{(l)} = \left(\mu_1^{(l)}, \mu_2^{(l)}, \mu_3^{(l)}, \mu_4^{(l)}, \tau_1^{(l)}, \tau_2^{(l)}, \tau_3^{(l)}, \tau_4^{(l)}, \psi^{(l)}, \beta^{(l)}\right).$$

对于每一个 l，一个牛肉热狗的卡路里数 $Y_1^{(l)}$ 服从均值为 $\mu_1^{(l)}$、方差为 $1/\tau_1^{(l)}$ 的正态分布. 同样一个家禽热狗的卡路里数服从均值为 $\mu_3^{(l)}$、方差为 $1/\tau_3^{(l)}$ 的正态分布. 然后计算 $D^{(l)} = Y_1^{(l)} - Y_3^{(l)}$. $D^{(1)}, \cdots, D^{(60\,000)}$ 值的样本分位数可以用于估计 D 的分布的分位数.

例如，假设想要得到 D 的 90% 的预测区间. 如上，我们模拟 60 000 个 $D^{(l)}$ 的值，找到 0.05 和 0.95 的样本分位数分别为 -14.86 和 87.35，即为预测区间的端点. 为了评价该模拟估计和 D 分布的实际分位数的接近程度，计算这两个端点的模拟标准误差. 对于 $k=6$ 个马尔可夫链中得到的样本，可以计算 D 值的 0.05 样本分位点. 然后可以把这些估计作为式(11.3.1)中的 $Z_1, \cdots, Z_6$ 去计算值 S. 模拟标准误差为 $S/6^{1/2}$. 对于 0.95 样本分位点，可以重复该程序. 对于区间的两个端点的模拟标准误差分别为 0.2447 和 0.3255. 相对于预测区间的长度而言这些模拟标准误差是非常小的. ◀

例 11.3.9 截尾的砷测量数据 Frey and Edwards(1997)描述了全国范围内砷分布情况的调查(NAOS). 几百个供水系统团体提交了他们的未经处理过的水样本，以期对全国的砷的分布特征描述提供帮助. 砷是环境保护机构(EPA)要求加以控制的几大污染物之一. 要构造一种像砷这种物质的发生情况的模型，其一大困难在于浓度经常太低以至于无法精确测量. 在此情况下，测量数据为截尾型的. 即，我们只知道砷的浓度低于某个截尾值，但是不知道低了多少. 在 NAOS 数据集中，截尾点是 0.5 微克/每升. 浓度低于 0.5 微克/每升的为截尾数据.

吉布斯抽样可以帮助我们估计带有截尾数据的砷的分布，Lockwood et al.(2001)对 NAOS 和其他数据做了广泛的分析，指出了不同州之间以及不同水资源之间砷的不同分布. 为方便起见，考察来自俄亥俄州的 24 个观察值. 在这 24 个观察值中，11 个是取自地下水资源(井)，另外 13 个是取自地表水资源(比如河流和湖泊). 下面是来自俄亥俄州的 7 个非截尾地下水观察值： [747]

9.62, 10.5, 2.30, 0.80, 17.04, 9.90, 1.32.

另外 4 个地下水观察值为截尾数据.

假设把俄亥俄州地下水中的砷浓度看作是参数为 μ 和 σ^2 的对数正态分布. 处理截尾数据的一个最普遍的方法是把它们看成是未知参数. 即，令 $Y_1, \cdots, Y_4$ 表示来

自四口井的未知浓度的截尾数据. 令 $X_1,\cdots,X_7$ 表示7个未截尾的值. 假设 μ 和 $\tau=1/\sigma^2$ 服从一个参数为 μ_0、λ_0、α_0 和 β_0 的正态-伽玛先验分布. $X_1,\cdots,X_7$, $Y_1,\cdots,Y_4$ 以及 μ 和 τ 的联合概率密度函数和下式成正比.

$$\tau^{\beta_0+(7+4+1)/2-1}\exp\left(-\frac{\tau}{2}\left[\lambda_0(\mu-\mu_0)^2+\sum_{i=1}^{7}(\ln(x_i)-\mu)^2+\sum_{j=1}^{4}(\ln(y_i)-\mu)^2+2\beta_0\right]\right).$$

观测的数据包括 $X_1,\cdots,X_7$ 的观察值 $x_1,\cdots,x_7$, 和 $Y_j\leqslant 0.5$, $j=1,\cdots,4$. 给定数据和其他参数下 μ 和 τ 的条件分布与在例11.3.1中得到的分布相同. 确切地说, 给定 τ、Y_j 以及数据的条件下, μ 服从正态分布, 其均值为

$$\frac{\lambda_0\mu_0+\sum_{i=1}^{7}x_i+\sum_{j=1}^{4}y_j}{\lambda_0+11},$$

精度为 $\tau(\lambda_0+11)$, 而 μ、Y_j 以及数据的条件下, τ 服从一个伽玛分布, 其参数为 $\alpha_0+(11+1)/2$ 和

$$\beta_0+\frac{1}{2}\left(\sum_{i=1}^{7}(x_i-\mu)^2+\sum_{j=1}^{4}(y_i-\mu)^2+\lambda_0(\mu-\mu_0)^2\right),$$

给定 μ、τ 以及数据的条件下, Y_j 为独立同分布的有对数正态分布的随机变量, 其参数为 μ 和 $1/\tau$, 但条件是 $Y_j<0.5$ 时, 即每个 Y_j 的条件分布函数是

$$F(y)=\frac{\Phi([\ln(y)-\mu]\tau^{1/2})}{\Phi([\ln(0.5)-\mu]\tau^{1/2})},\quad 当\ y<0.5\ 时.$$

只要计算出标准正态分布的分布函数和分位数函数, 就可以模拟出分布函数为 F 的随机变量. 设 U 服从区间[0, 1]上的均匀分布, 则

$$Y=\mu+\tau^{-1/2}\Phi^{-1}[U\Phi([\ln(0.5)-\mu]\tau^{1/2})]$$

748

的分布函数即为 F.

在此类分析中需要做推断的一个例子是预测不同水系统中的砷浓度. 知道砷测量的可能值可以帮助水系统选择比较经济的处理方式, 来达到环境保护机构的标准. 为简单起见, 可以在马尔可夫链的每次迭代中模拟一种砷浓度. 比如说, 假设 $(\mu^{(i)},\tau^{(i)})$ 是在马尔可夫链的第 i 次迭代中 μ 和 τ 的模拟值, 那么可以模拟 $Y^{(i)}=\exp(\mu^{(i)}+Z(\tau^{(i)})^{-1/2})$, 其中 Z 是一个标准正态随机变量. 图11-8给出了从每个长度为10 000的10个马尔可夫链中得到的模拟值 $\ln(Y^{(i)})$ 的直方图. $\ln(0.5)$ 以下的截尾值的比例是0.335, 模拟标准误差是0.007. 对数刻度下的预测值的中位数为0.208, 用例11.1.8中描述的 δ 方法可将预测值转换为原来的尺度. 砷浓度的预测值的中位数是 $\exp(0.208)=1.231$ 微克/升. 模拟标准误差为 $0.007\exp(0.208)=0.009$. ◀

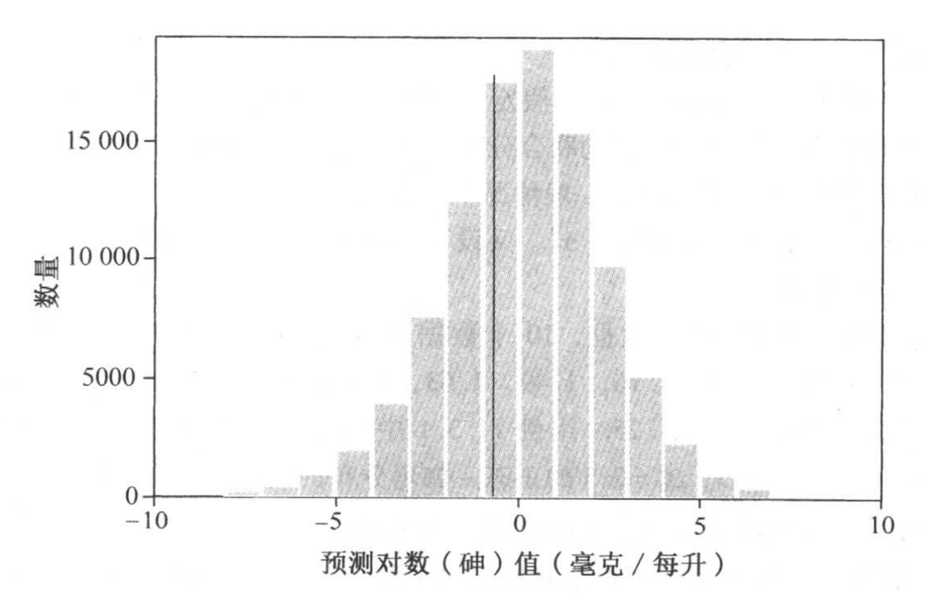


图 11-8 例 11.3.9 中对 10 个马尔可夫链中的每个进行 10 000 次迭代的模拟对数(砷)值的直方图．垂直线的位置是在截尾点 ln(0.5)处

注：有更一般的马尔可夫链蒙特卡罗法．对于要模拟的分布，吉布斯抽样要求有特殊的结构，我们要能在给定其他坐标下模拟出每一坐标的分布．在许多问题中，这是不可能的，即使不是对全部坐标，至少对有些坐标是不行的．如果只有一个坐标较难模拟，可以尝试对该坐标使用接受/拒绝模拟．若还行不通，还可以用更一般的马氏链蒙特卡罗法．其中最简单的是由 Metropolis et al.（1953）提出的 Metropolis 法．在 Gelman et al.（1953）编写的书的第 11 章有 Meropolis 法的介绍，还有 Hastings（1970）提出的更一般的方法．

749

11.3.7 小结

我们介绍了从感兴趣的联合分布产生观测值的马尔可夫链的吉布斯抽样算法．该联合分布必须有一个特殊的形式．作为每一个变量的函数，联合概率密度必须看成一个概率密度函数，由此易于模拟出伪随机变量．吉布斯抽样算法通过坐标进行循环，在其余变量给定下对每个坐标进行模拟，算法需要一个曝光期，在此期间马尔可夫链的状态分布会收敛到想要的分布．通过同步地运行几个独立的马尔可夫链，可以评估收敛性并计算模拟值的模拟标准误差．

习题

1. 设 $f(x_1,x_2)=cg(x_1,x_2)$ 为 (X_1,X_2) 的联合密度函数．对于每个给定的 x_2，令 $h_2(x_1)=g(x_1,x_2)$．即当 x_2 固定时，h_2 是 $g(x_1,x_2)$ 关于 x_1 的函数．证明：必存在不依赖于 x_1 的因子 c_2，使得 $h_2(x_1)c_2$ 是 X_1 在 $X_2=x_2$ 条件下的条件密度函数．
2. 设 $f(x_1,x_2)$ 是联合密度函数，假设 $(X_1^{(i)},X_2^{(i)})$ 有联合密度函数 f. 应用本节给出的吉布斯抽样法的第 2 和第 3 步，将得到的结果记作 $(X_1^{(i+1)},X_2^{(i+1)})$．证明：$(X_1^{(i+1)},X_2^{(i)})$ 和

$(X_1^{(i+1)}, X_2^{(i+1)})$也有联合密度函数$f$.

3. 设$Z_1, Z_2, \cdots$形成一个马尔可夫链，假定Z_1的分布是平稳分布. 证明：对于任意$i>1$，(Z_1, Z_2)的联合分布与(Z_i, Z_{i+1})的联合分布一致. 为了简单起见，可以假定马尔可夫链有有限的状态空间，不过其结果在一般情况下也成立.
4. 设$X_1, \cdots, X_n$互不相关，方差均为σ^2. 又设$Y_1, \cdots, Y_n$正相关，方差均为σ^2. 证明：$\overline{X}$的方差小于$\overline{Y}$的方差.
5. 使用奶酪乳酸含量的30个数据，10个数据来自例7.5.2，20个数据为：1.68，1.9，1.06，1.3，1.52，1.74，1.16，1.49，1.63，1.99，1.15，1.33，1.44，2.01，1.31，1.46，1.72，1.25，1.08，1.25. 用例11.3.1中的吉布斯抽样法拟合一个先验分布为$\mu_0=1$，$\lambda_0=1$，$\alpha_0=0.5$，$\beta_0=0.5$的正态-伽玛分布的模型. 模拟10 000组(μ, τ)参数. 估计$(\sqrt{\tau}\mu)^{-1}$的后验均值，计算估计值的模拟标准误差.
6. 洗碗机的数据如表11-6所示. 假定我们根据时间（年份减去1960）和私人住宅投资对预测洗衣机的销售量拟合一个多元线性回归模型. 假设参数有一个不合适的先验分布，它与$1/\tau$成比例. 运用吉布斯抽样，从参数的联合后验分布中得到一个容量为10 000的样本.

表11-6　1960~1985期间的洗碗机装运和住宅投资

年份	洗碗机装运（千）	私人住宅投资（十亿美元）
1960	555	34.2
1961	620	34.3
1962	720	37.7
1963	880	42.5
1964	1050	43.1
1965	1290	42.7
1966	1528	38.2
1967	1586	37.1
1968	1960	43.1
1969	2118	43.6
1970	2116	41.0
1971	2477	53.7
1972	3199	63.8
1973	3702	62.3
1974	3320	48.2
1975	2702	42.2
1976	3140	51.2
1977	3356	60.7
1978	3558	62.4
1979	3488	59.1
1980	2738	47.1
1981	2484	44.7
1982	2170	37.8
1983	3092	52.7
1984	3491	60.3
1985	3536	61.4

a. 设 β_1 为时间的系数. 用你所得到的后验样本绘制 $|\beta_1|$ 的样本密度函数.

b. 我们对预测 1986 年洗碗机销售感兴趣.

i. 从后验分布画出 $\beta_0+26\beta_1+67.2\beta_2$ 值的直方图.

ii. 对于每个参数的模拟值，模拟出 1986 年洗碗机的销售量(时间 = 26, 私宅投资 = 67.2). 计算模拟值计算的置信度为 90% 的预测区间.

iii. 绘制 1986 年洗碗机的模拟销售量的直方图，与 i. 中的直方图做比较. 请解释为什么其中一个样本与其他比较起来方差较大.

7. 用 10.3 节习题 10 表中的数据. 这次拟合例 11.3.7 中的模型. 用先验超参数 $\lambda_0=\alpha_0=1$, $\beta_0=0.1$, $u_0=0.001$ 和 $\psi_0=800$. 从参数的后验联合分布中得到一个容量为 10 000 的样本. 估计三个参数 μ_1, μ_2, μ_3 的后验均值.

750

8. 本问题中，我们对稳健(robust)线性回归做个概述. 假设本习题中的数据都以 (Y_i, x_i), $i=1,\cdots,n$ 的形式成对出现. 假设所有 x_i 均已知，Y_i 均为独立的随机变量. 此处只考虑简单的回归，不过该方法可以很容易推广到多元回归.

a. 设 β_0, β_1, σ 代表未知参数，令 a 为一个已知的正常数. 证明：以下两个模型等价，即证明在两个模型中 $(Y_1,\cdots,Y_n)$ 的联合分布相同.

模型 1：对于每个 i, $[Y_i-(\beta_0+\beta_1x_i)]/\sigma$ 服从一个自由度为 a 的 t 分布.

模型 2：对于每个 i, 给定 τ_i 时 Y_i 服从均值为 $\beta_0+\beta_1x_i$、方差为 $1/\tau_i$ 的条件正态分布. 并且 $\tau_1,\cdots,\tau_n$ 独立同分布，服从参数为 $a/2$ 和 $a\sigma^2/2$ 的伽玛分布.

提示：当 μ 和 τ 服从正态伽玛分布时，设法产生 μ 的边际分布.

b. 现在考虑 a. 中的模型 2，令 $\eta=\sigma^2$，假定 η 服从参数为 $b/2$ 和 $f/2$ 的先验伽玛分布，其中 b 和 f 为已知常量. 假定参数 β_0 和 β_1 有一个不恰当的联合先验，其概率密度函数为 1. 证明：似然函数和先验联合密度函数的乘积是一个常数乘以

$$\eta^{(na+b)/2-1}\prod_{i=1}^{n}\tau_i^{(a+1)/2-1}\exp\left(-\frac{1}{2}\left[f\eta+\sum_{i=1}^{n}\tau_i\{a\eta+(y_i-\beta_0-\beta_1x_i)^2\}\right]\right). \quad (11.3.4)$$

c. 考虑(11.3.4)是每一个参数在其他参数为固定值下的函数. 证明：表 11-7 给出了每个参数在已知其他条件下的恰当的条件分布.

表 11-7 习题 8 中的参数和条件分布

参数	(11.3.4)看似为以下这个分布的密度函数
η	参数为 $(na+b)/2$ 和 $(f+a\sum_{i=1}^{n}\tau_i)/2$ 的伽玛分布
τ_i	参数为 $(a+1)/2$ 和 $[a\eta+(y_i-\beta_0-\beta_1x_i)^2]/2$ 的伽玛分布
β_0	均值为 $\sum_{i=1}^{n}\tau_i(y_i-\beta_1x_i)/\sum_{i=1}^{n}\tau_i$ 和精度为 $\sum_{i=1}^{n}\tau_i$ 的正态分布
β_1	均值为 $\sum_{i=1}^{n}\tau_ix_i(y_i-\beta_0)/\sum_{i=1}^{n}\tau_ix_i^2$ 和精度为 $\sum_{i=1}^{n}\tau_ix_i^2$ 的正态分布

9. 使用 10.2 节的表 10-5 的数据，假设 Y 是压力的对数，x_i 是第 i 次观测到的沸点，$i=1,\cdots,17$，采用习题 8 中的稳健(robust)回归方法，其中 $a=5, b=0.1, f=0.1$. 估计参数

β_0，β_1 和 η 的后验均值和标准离差.

10. 本问题中，我们将概述6.3节例6.4.8所描述的问题的贝叶斯解. 令 $\tau=1/\sigma^2$，除了两个参数 μ 和 τ 外，引入 n 个附加参数. 对 $i=1,\cdots,n$，如果 X_i 来自于均值为 μ 精度为 τ 的正态分布，则令 $Y_i=1$；如果 X_i 来自于标准正态分布，则 $Y_i=0$.
    a. 给定 τ、$Y_1,\cdots,Y_n$ 和 $X_1,\cdots,X_n$，求 μ 的条件分布.
    b. 给定 μ、$Y_1,\cdots,Y_n$ 和 $X_1,\cdots,X_n$，求 μ 的条件分布.
    c. 给定 μ、τ、$X_1,\cdots,X_n$ 和其他的 Y_j，求 Y_j 的条件分布.
    d. 叙述怎样用吉布斯抽样找到 μ 和 τ 的后验分布.
    e. 证明：Y_i 的后验均值为 X_i 来自于均值和方差均未知的正态分布的后验概率.
11. 再次考虑例6.4.8中所描述的模型. 假设 $n=10$，$X_1,\cdots,X_{10}$ 的观测值是
    -0.92，-0.33，-0.09，0.27，0.50，-0.60，1.66，-1.86，3.29，2.30
    a. 使用习题10中的吉布斯抽样算法将模型与观测数据作拟合，用以下的先验超参数：$\alpha_0=1$，$\beta_0=1$，$\mu_0=0$，$\lambda_0=1$.
    b. 对于每个 i 估计 X_i 来自于未知均值和方差的正态分布的后验概率.

751

12. 令 $X_1,\cdots,X_n$ 是独立同分布，都有一个均值为 μ、精度为 τ 的正态分布，吉布斯抽样允许对 (μ,τ) 使用一个先验分布，其中 μ 和 τ 是独立的. 令 μ 的先验分布为一个均值为 μ_0、方差为 γ_0 的正态分布. 令 τ 的先验分布为以 α_0 和 β_0 为参数的伽玛分布.
    a. 证明表11-8给出了给定其他的条件下每个参数的合适的条件分布.
    b. 用新墨西哥护理院的数据. 设先验的超参数为 $\alpha_0=2,\beta_0=6300,\mu_0=200,\gamma_0=6.35\times10^{-4}$. 用吉布斯抽样找到 (μ,τ) 的后验分布. 特别地，计算一个包含95%的 μ 的后验分布的区间.

表11-8 习题12的参数和条件分布

参数	先验乘以似然分布的概率密度函数
τ	参数为 $\alpha_0+n/2$ 和 $\beta_0+0.5\sum_{i=1}^{n}(x_i-\bar{x})^2+0.5n(\bar{x}-\mu)^2$ 的伽玛分布
μ	均值为 $(\gamma_0\mu_0+n\tau\bar{x})/(\gamma_0+n\tau)$ 精度为 $\gamma_0+n\tau$ 的正态分布

13. 再次考虑习题12中所描述的情形，此次，我们令 μ 的先验分布更像是个共轭先验. 引进另一个参数 γ，它的先验分布是一个参数为 a_0 和 b_0 的伽玛分布. 令 μ 在给定 γ 条件下的先验分布服从于一个均值为 μ_0 和精度为 γ 的正态分布.
    a. 证明：由 μ 的边际先验分布确定 $\left(\dfrac{b_0}{a_0}\right)^{1/2}(\mu-\mu_0)$ 服从自由度为 $2a_0$ 的 t 分布.
    b. 假设我们希望 μ 和 τ 的边际先验分布和它们的共轭先验分布相同. 为此，先验超参数之间必须有什么联系？
    c. 证明：表11-9给出了给定其他参数下每个参数的合适的条件分布.
    d. 使用新墨西哥护理院的数据（例11.3.3和例11.3.4）. 令先验超参数为 $\alpha_0=2$，$\beta_0=6300$，$\mu_0=200$，$a_0=2$，$b_0=3150$. 实施一个吉布斯抽样以得到 (μ,τ,γ) 的后验分布. 特别地，计算包含95%的 μ 的后验分布的区间.

表 11-9 习题 13 的参数和条件分布

参数	先验乘以似然分布的概率密度函数
τ	参数为 $\alpha_0+n/2$ 和 $\beta_0+0.5\sum_{i=1}^{n}(x_i-\bar{x})^2+0.5n(\bar{x}-\mu)^2$ 的伽玛分布
μ	均值为 $(\gamma\mu_0+n\tau\bar{x})/(\gamma+n\tau)$ 和精度为 $\gamma+n\tau$ 的正态分布
γ	参数为 $a_0+1/2$ 和 $b_0+0.5\ (\mu-\mu_0)^2$ 的伽玛分布

14. 考虑例 11.3.9 所描述的情形，除了 11 个地下水源外，其他还有 13 个观测值来自于俄亥俄州的地表水资源. 在这 13 个地表水测量中，只有一个是截尾的. 从俄亥俄州得到的没有被截尾的地表水含砷的浓度为

    1.93，0.99，2.21，2.29，1.15，1.81，2.26，3.10，1.18，1.00，2.67，2.15

    a. 拟合与例 11.3.9 所描述的相同的模型，并预测对马尔可夫链的每次迭代计算地表水浓度的对数.

    b. 比较你所预测的测量值的直方图与图 11-8 的地下水预测值的直方图. 描述主要的差别.

    c. 估计地表水的砷浓度分布的中位数并与被预测的地下水浓度分布的中位数做比较.

752

15. 令 $X_1,\cdots,X_{n+m}$ 为一个来自于参数为 θ 的指数分布的随机样本. 假设 θ 服从于有已知参数 α 和 β 的伽玛先验分布. 假设我们观测到 $X_1,\cdots,X_n$，但 X_{n+1}，…，X_{n+m} 是截尾的.

    a. 首先，假设截尾工作是如下进行的. 对于 $i=1,\cdots,m$，如果 $X_{n+i}\leqslant c$，我们只知道 $X_{n+i}\leqslant c$，但不是 X_{n+i} 的精确值. 建立一个吉布斯抽样算法，使我们能模拟 θ 的后验分布，尽管有的观测被截尾.

    b. 接下来，假设截尾工作如下. 对于 $i=1,\cdots,m$，如果 $X_{n+i}\geqslant c$，我们只知道 $X_{n+i}\geqslant c$，但不知道 X_{n+i} 的精确值. 建立一个吉布斯抽样算法，以使我们能模拟 θ 的后验分布，尽管有的观测被截尾.

16. 假设完成某任务的时间是 X 和 Y 两部分的总和. 令 (X_i,Y_i)（其中 $i=1,\cdots,n$）为完成两部分任务的时间的一个随机样本. 但是，对于一些观察值，我们观察到的只是 $Z_i=X_i+Y_i$. 为了确切些，假设对于 $i=1,\cdots,k$，我们观察到了 (X_i,Y_i)，对于 $i=k+1,\cdots,n$，我们观察到了 Z_i. 假设所有 X_i 和 Y_i 是独立的，并且每个 X_i 服从于一个参数为 λ 的指数分布，每个 Y_i 服从于一个参数为 μ 的指数分布.

    a. 证明：给定 $Z_i=z$ 时 X_i 的条件分布的密度函数为

    $$G(x\mid z)=\frac{1-\exp(-x[\lambda-\mu])}{1-\exp(-z[\lambda-\mu])},\text{其中 }0<x<z.$$

    b. 假设 (λ,μ) 的先验分布如下：两个参数是独立的，并且 λ 服从于参数为 a 和 b 的伽玛分布，且 μ 服从于参数为 c 和 d 的伽玛分布. 这 4 个数字 a，b，c，d 都是已知的常数. 建立一个吉布斯抽样法来模拟 (λ,μ) 的后验分布.

11.4 自 助 法

参数和非参数的自助法是一种在概率计算中用已知分布来代替未知分布 F 的方

法. 如果我们得到一组来自于分布 F 的数据样本，首先用 $\hat{F}$ 来近似 F，然后进行所要做的计算. 如果 $\hat{F}$ 是 F 的一个很好的近似，那么自助法就成功了. 如果要做的计算相当复杂，则可以用模拟的方法.

11.4.1 引言

假定我们有一组来自未知分布 F 的数据样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$，假设我们对于涉及 F 和 $\boldsymbol{X}$ 的某个量感兴趣，比如，把统计量 $g(\boldsymbol{X})$ 的偏差作为 F 中位数的估计. 在这个最简单的例子中自助分析背后的主要思想是：第一步，用一个已知分布 $\hat{F}$ 来代替未知分布 F. 接着，令 $\boldsymbol{X}^*$ 是来自于分布 $\hat{F}$ 的样本. 最后，基于 $\hat{F}$ 和 $\boldsymbol{X}^*$ 来计算我们所感兴趣的量. 例如，用 $g(\boldsymbol{X}^*)$ 的偏差作为 $\hat{F}$ 的中位数的估计值. 考虑下面一个简单的例子.

例 11.4.1 样本均值的方差 令 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从有连续分布函数 F 的分布中得到的一个随机样本，现在我们只假定 F 有有限的均值 μ 和有限的方差 σ^2. 如果对于样本均值 $\overline{X}$ 的方差感兴趣，我们已经知道该方差等于 σ^2/n，但是不知道 σ^2 等于多少. 为了估计 σ^2/n，自助法用已知分布 $\hat{F}$ 来代替未知分布 F，$\hat{F}$ 同样也有有限的均值 μ 和有限的方差 σ^2. 如果 $\boldsymbol{X}^*=(X_1^*,\cdots,X_n^*)$ 是来自于 $\hat{F}$ 的一个随机样本，那么该样本均值 $\overline{X}^*=\frac{1}{n}\sum_{i=1}^{n}X_i^*$ 的方差是 σ^2/n. 既然 $\hat{F}$ 是已知的，就应该可以计算出 $\hat{\sigma}^2/n$，于是可以用此值来估计 σ^2/n 的值.

753

对已知分布 $\hat{F}$ 通常的选择是样本分布函数 F_n. 该样本分布函数 F_n 是离散的分布函数，在随机样本 $X_1,\cdots,X_n$ 的每个观察值 $x_1,\cdots,x_n$ 上都有幅度为 $1/n$ 的跳跃，所以，如果 $\boldsymbol{X}^*=(X_1^*,\cdots,X_n^*)$ 是 $\hat{F}$ 的随机样本，每个 X_i^* 是离散随机变量，它的概率密度函数为：

$$f(x)=\begin{cases}\dfrac{1}{n} & 若\ x\in\{x_1,\cdots,x_n\},\\ 0 & 其他.\end{cases}$$

因此使得计算密度函数为 f 的随机变量 X_i^* 的方差变得的 $\hat{\sigma}^2$ 相对简单. 该方差为：

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

其中 $\bar{x}$ 是观察值 $x_1,\cdots,x_n$ 的均值. 这样，我们对 $\overline{X}$ 的方差的自助估计为 $\hat{\sigma}^2/n$. —◀

在自助法分析中关键的步骤就是对于已知分布 $\hat{F}$ 的选择. 例 11.4.1 作的特殊的选择是样本分布函数，此种自助法称为非参数自助法. 这是因为当选择 $\hat{F}=F_n$ 时，并没有假定分布 F 属于一个参数系. 若我们愿意假设 F 属于某个参数系，则可以选择该参数系中的 $\hat{F}$，然后利用参数自助分析，下面举例子说明.

例 11.4.2 样本均值的方差 令随机样本 $X=(X_1,\cdots,X_n)$ 为来自于均值为 μ、方差为 σ^2 的正态分布. 假定与例 11.4.1 中一样，要估计样本均值 $\overline{X}$ 的方差 σ^2/n. 运用参数自助法，用正态分布族中的一个 $\hat{F}$ 来代替 F. 在本例中，$\hat{F}$ 为正态分布，虽然还有其他的选择，我们选择 $\hat{F}$ 为正态分布，其均值和方差分别等于极大似然估计 $\bar{x}$ 和 $\hat{\sigma}^2$. 然后用来自于分布 $\hat{F}$ 的随机样本的样本均值 $\overline{X}^*$ 的方差来估计 σ^2/n. $\overline{X}^*$ 的方差很容易计算，即为 $\hat{\sigma}^2/n$. 本例中，用参数自助法得到了和非参数自助法几乎相同的结果. ◀

在例 11.4.1 和例 11.4.2 中，很容易计算取自分布 $\hat{F}$ 的随机样本的样本均值的方差. 在自助法的典型应用中，计算感兴趣的量并不是很容易. 例如，在例 11.4.1 和例 11.4.2 中，没有简单的公式可以计算 $\hat{F}$ 的样本 X^* 的样本中位数的方差，在此情况下，可以借助于模拟方法来近似所需的计算. 在举例说明关于如何在自助法中运用模拟的方法之前，我们将先描述使用自助分析的一般情形. 754

11.4.2 一般的自助法

设 $\eta(X,F)$ 是感兴趣的量，它可能依赖于 F 分布及其样本 X. 例如若 F 分布具有概率密度函数 f，我们或许对下面的量感兴趣：

$$\eta(X,F)=\left[\frac{1}{n}\sum_{i=1}^{n}X_i-\int xf(x)\mathrm{d}x\right]^2. \tag{11.4.1}$$

在例 11.4.1 和例 11.4.2 中，我们要求得到样本均值的方差，它等于(11.4.1)式的均值. 通常，我们希望估计 $\eta(X,F)$ 的均值、分位数或其他一些概率特征. 在自助法中，用 $\eta(X^*,\hat{F})$ 的均值、分位数或其他一些概率特征来估计 $\eta(X,F)$ 的均值、分位数或另外一些概率特征，其中 X^* 是从 $\hat{F}$ 分布抽取的随机样本，$\hat{F}$ 是某个我们希望接近 F 的分布. 表 11-10 显示了数据的原始统计模型和用自助分析得到的量之间的对应关系. 感兴趣的函数 η 必须对于所考虑的所有分布和来自于这些分布的样本都存在的一些量. 其他感兴趣的量可能是某统计量的分布的分位数，某个估计的 M. A. E. 或 M. S. E.，某个估计的偏差，统计量属于某些区间的概率等.

表 11-10 统计模型和自助分析的相似性

	统计模型	自助法
分布	F 未知	已知 $\hat{F}$
数据	来自 F 的独立同分布的样本 X	来自 $\hat{F}$ 独立同分布的样本 X^*
感兴趣的函数	$\eta(X,F)$	$\eta(X^*,\hat{F})$
参数/估计	$\eta(X,F)$ 的均值、中位数、方差等	$\eta(X^*,\hat{F})$ 的均值、中位数、方差等

在此前遇到的那些简单例子中，$\eta(X^*,\hat{F})$的分布是已知的，而且是容易计算的.通常情况下，$\eta(X^*,\hat{F})$的分布过于复杂而不能解析地计算它的特征. 此时可用下面的模拟方法近似自助估计. 第一步，从分布 $\hat{F}$ 中抽取大量(记为 ν 个)的随机样本 $X^{*(1)},\cdots,X^{*(\nu)}$. 然后计算 $T^{(i)}=\eta(X^{*(i)},\hat{F})$，$i=1,\cdots,\nu$. 最后计算所要的 $T^{(1)},\cdots,T^{(\nu)}$的样本分布函数的特征.

例 11.4.3 样本中位数的极大似然估计 假设要对来自于中位数为 θ 的分布函数 F 的连续分布的数据 $X=(X_1,\cdots,X_n)$建立模型，又设我们对于用样本中位数 M 作为 θ 的估计有兴趣. 要估计 M 的 M. S. E 作为 θ 的估计，即令 $\eta(X,F)=(M-\theta)^2$，要估计 $\eta(X,F)$的均值. 令 $\hat{F}$ 为一个已知的且类似于 F 的分布，X^* 为来自于 $\hat{F}$ 的样本容量为 n 的随机样本. 不管选择哪个分布 $\hat{F}$，都很难计算自助估计，即 $\eta(X^*,\hat{F})$的均值. 取而代之，我们模拟大数量为 ν 的来自于 $\hat{F}$ 分布的样本 $X^{*(1)},\cdots,X^{*(\nu)}$，计算每个样本的样本中位数 $M^{(1)},\cdots,M^{(\nu)}$. 然后计算 $T^{(i)}=(M^{(i)}-\hat{\theta})^2$，$i=1,\cdots,\nu$，其中 $\hat{\theta}$是 $\hat{F}$ 分布的中位数. 我们模拟近似的自助估计结果就是 $T^{(1)},\cdots,T^{(\nu)}$的均值.

755

举个例子，设有一个样本容量 $n=25$ 的样本，样本值为 $y_1,\cdots,y_{25}$在表 11-11 中给出.

表 11-11 Kolmogorov-Smirnov 检验的计算

i	y_i	$F_n(y_i)$	$\Phi(y_i)$
1	-2.46	0.04	0.0069
2	-2.11	0.08	0.0174
3	-1.23	0.12	0.1093
4	-0.99	0.16	0.1611
5	-0.42	0.20	0.3372
6	-0.39	0.24	0.3483
7	-0.21	0.28	0.4168
8	-0.15	0.32	0.4404
9	-0.10	0.36	0.4602
10	-0.07	0.40	0.4721
11	-0.02	0.44	0.4920
12	0.27	0.48	0.6064
13	0.40	0.52	0.6554
14	0.42	0.56	0.6628
15	0.44	0.60	0.6700
16	0.70	0.64	0.7580
17	0.81	0.68	0.7910
18	0.88	0.72	0.8106
19	1.07	0.76	0.8577
20	1.39	0.80	0.9177
21	1.40	0.84	0.9192
22	1.47	0.88	0.9292
23	1.62	0.92	0.9474
24	1.64	0.96	0.9495
25	1.76	1.00	0.9608

对非参数的自助分析，使用 $\hat{F}=F_n$. 注意分布 $\hat{F}$ 的中位数就是原始样本的样本中位数 $\hat{\theta}=0.40$. 然后，我们从分布 $\hat{F}$ 中抽取 $v=10\ 000$ 个容量为 25 的随机样本. 即从 y_i 个值中有放回地抽取 25 个数，重复抽取 10 000 个容量为 25 的样本(解习题 2 以证明为什么这样做可得到所需的样本 $X^{*(1)},\cdots,X^{*(v)}$). 例如，这是 10 000 个自助样本中的一个：

1.64	0.88	0.70	−1.23	−0.15	1.40	−0.07	−2.46	−2.46	−0.10
−0.15	1.62	0.27	0.44	−0.42	−2.46	1.40	−0.10	0.88	0.44
−1.23	1.07	0.81	−0.02	1.62					

如果把样本中的数排序，可得样本中位数为 0.27. 事实上，10 000 个样本中有 1405 个自助样本的中位数等于 0.27. 图 11-9 给出了从自助取样的所有 10 000 个样本中位数的直方图. 原始样本中四个最大以及最小的观察值从未在 10 000 个自助样本中作为样本中位数出现. 对于 10 000 个自助样本中每个 i，计算样本中位数 $M^{(i)}$ 以及平方误差 $T^{(i)}=(M^{(i)}-\hat{\theta})^2$，其中 $\hat{\theta}=0.40$ 是 $\hat{F}$ 分布的中位数，然后对 10 000 个样本的所有值取平均值，得到 0.0887. 即为我们对样本中位数的 M. S. E. 的非参数自助估计的模拟近似. 模拟值 $T^{(i)}$ 的样本方差为 $\hat{\sigma}^2=0.0135$，自助估计的模拟标准差为 $\hat{\sigma}/\sqrt{10\ 000}=1.163\times10^{-3}$. ◀

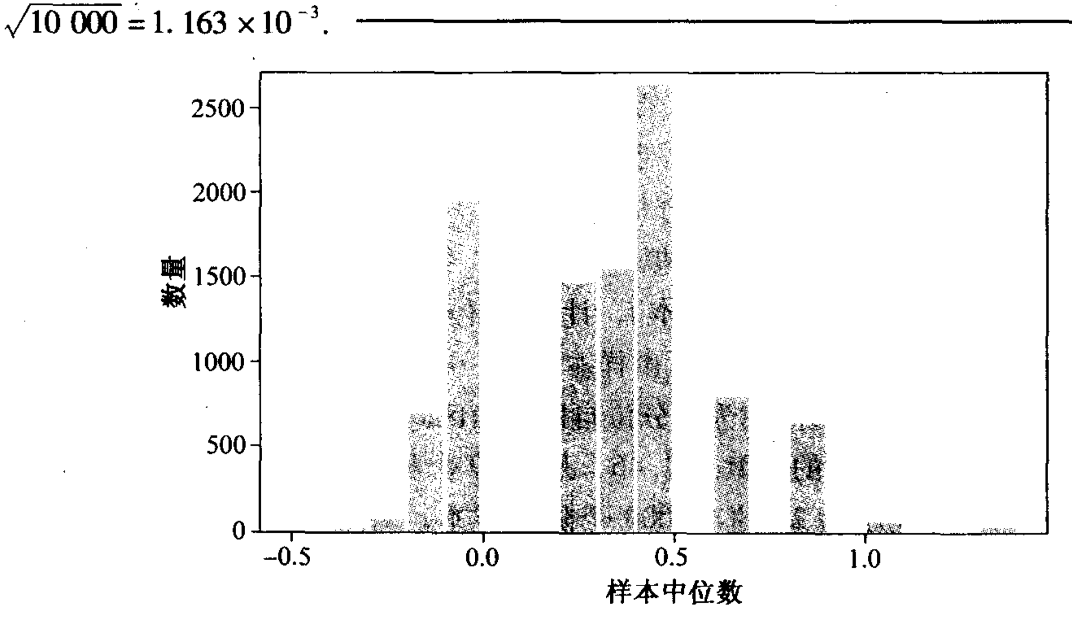


图 11-9 例题 11.4.3 中 10 000 个自助样本的样本中位数

注： *自助估计的模拟近似法.* 自助法是一种估计技术. 它产生我们感兴趣的参数的估计. 当自助估计太难计算时，就用模拟的方法，模拟能够提供自助估计的估计值. 本书中将自助估计的模拟估计值看作近似值. 我们这样做仅仅是为了避免叫做估计的估计量.

自助法是由 Efron(1979)提出的，至今已有许多的应用. 读者若要了解更多关于自助法的细节可参阅 Efron 和 Tibshirani(1993)或 Davison 和 Hinkley(1997). Young(1994)

756 回顾了很多关于自助法的论著并包含了很多有用的文献. 在本章接下来的部分，我们将给出一些有关参数和非参数自助法的例子，并解释如何使用模拟的方法来近似所要的自助估计.

11.4.3 非参数自助法

例 11.4.4 四分位距的置信区间 四分位距(IQR)的定义为上侧四分位和下侧四分位的差距，即0.75和0.25分位数之间的差距. 分布的中心50%的比例落于上下四分位数之间. 所以IQR是包含了分布中间二分之一的区间的长度. 例如：如果F是一个方差为σ^2的正态分布，则它的IQR为1.35σ.

若欲求未知分布F的IQRθ的90%置信区间，我们有来自于分布F的随机样本$X_1,\cdots,X_n$. 构造置信区间的方法很多，因此只考虑那些建立在θ和样本IQR$\hat{\theta}$之间关系基础上的方法. 由于IQR是一个尺度特性，因此在$\hat{\theta}/\theta$的分布上建立置信区间是合理的. 即令$\hat{\theta}/\theta$的分布的0.05和0.95分位数为a和b，则有

$$\Pr\left(a\leqslant\frac{\hat{\theta}}{\theta}\leqslant b\right)=0.9.$$

由于$a\leqslant\hat{\theta}/\theta\leqslant b$与$\hat{\theta}/b\leqslant\theta\leqslant\hat{\theta}/a$等价，可以得到结论$(\hat{\theta}/b,\hat{\theta}/a)$为$\theta$的90%的置信区间. 下面介绍非参数自助法对于分位数a和b的估计. 令$\eta(\boldsymbol{X},F)=\hat{\theta}/\theta$为样本$\boldsymbol{X}$

757 的样本IQR与分布F的IQR的比率. 令$\hat{F}=F_n$. 注意，$\hat{F}$的IQR为$\hat{\theta}$，即样本IQR. 再令$\boldsymbol{X}^*$为来自$\hat{F}$容量为n的样本，$\hat{\theta}^*$为由X^*计算出的样本IQR. 于是$\eta(\boldsymbol{X}^*,\hat{F})=\hat{\theta}^*/\hat{\theta}$. $\eta(X,F)$分布的0.05和0.95分位数由$\eta(\boldsymbol{X}^*,\hat{F})$的分布的0.05和0.95分位数估计. 后面这些分位数是通过模拟来近似的. 我们模拟大数量v的自助样本$\boldsymbol{X}^{*(i)},i=1,\cdots,v$. 对于每一个自助样本$i$，计算出样本的IQR$\hat{\theta}^{*(i)}$，然后用它除以$\hat{\theta}$，称该比率为$T^{(i)}$，$\hat{\theta}^*/\hat{\theta}$的$q$分位数可由样本$T^{(1)},\cdots,T^{(v)}$的样本$q$分位数近似得出. 用该种方法构造出的置信区间称为百分位自助置信区间.

现举例说明，分布F_n的IQR为1.46，是第19个和第6个观察值的差. 我们模拟出来自分布F_n的大小为25的10 000个随机样本，对第i个样本，计算出样本IQR$\hat{\theta}^{*(i)}$然后除以1.46，得到$T^{(i)}$，把$T^{(1)},\cdots,T^{(10\,000)}$从小到大排序后的第500个和第9500个的值为0.5822和1.6301，然后计算出非参数自助置信区间为$(1.46/1.6301,1.46/0.5822)=(0.8956,2.5077)$. ◀

例 11.4.5 位置参数的置信区间 令$X_1,\cdots,X_n$为来自于分布F的一个随机样本. 假设要求F的中位数的置信区间，可以在样本中位数M基础上建立一个置信区间. 例如，区间的形式可能为$[M-c_1,\ M+c_2]$. 由于$M-c_1\leqslant\theta\leqslant M+c_2$等价于$-c_2\leqslant M-\theta\leqslant c_1$，我们可以把$-c_2$和$c_1$看作$M-\theta$的分布的分位数. 如果不对$F$分布做假设，要得到$M-\theta$的分布的近似分位数是很难的. 为了计算百分位自助置信区间，

令 $\eta(\boldsymbol{X},F)=M-\theta$，然后将分布 $\eta(\boldsymbol{X},F)$ 的分位数（如 $\alpha_0/2$ 与 $1-\alpha_0/2$）用对应的 $\eta(\boldsymbol{X}^*,\hat{F})$ 的分位数来近似．此处 $\hat{F}$ 是样本的分布函数 F_n，其中位数为 M，$\boldsymbol{X}^*$ 为来自 $\hat{F}$ 的随机样本．然后选择一个比较大的数 v，模拟出很多样本 $\boldsymbol{X}^{*(i)}$，$i=1,\cdots,v$．对每一个样本，计算样本中位数 $M^{*(i)}$，再找出值 $M^{*(i)}-M$，$i=1,\cdots,v$ 的样本分位数．◄

例 11.4.5 中百分位自助置信区间的好坏关键在于 M^*-M 的分布对 $M-\theta$ 的分布近似的程度（此处 M^* 是来自分布 $\hat{F}$ 的容量为 n 的样本 $\boldsymbol{X}^*$ 的中位数）．例 11.4.5 中的近似值是可以改进的．M^*-M 的分布与 $M-\theta$ 的分布有差异的原因之一是其中一个分布或多或少地比另一个分布更分散．我们可以使用另一个分散度的差异尽可能小的自助近似．可将形如 $[M-c_1,\ M-c_2]$ 的区间用区间 $[M-d_1Y,\ M-d_2Y]$ 代替，其中 Y 为测量数据分散度的统计量．Y 可以为样本 IQR．另一种可能的分散性度量为样本中位数绝对偏差（即 $|X_1-M|,\cdots,|X_n-M|$ 的样本中位数）．现在我们看到 $M-d_1Y\leqslant\theta\leqslant M-d_2Y$ 等价于

$$-d_2\leqslant\frac{M-\theta}{Y}\leqslant d_1.$$

758

于是希望 $-d_2$ 和 d_1 是 $(M-\theta)/Y$ 的分布的分位数．该区间的形式类似于在 7.5 节中构建的 t 置信区间．的确，称所构造的该区间为 t-百分位数自助置信区间．我们使用如下的自助样本 $\boldsymbol{X}^*$ 构造 t-百分位数自助置信区间．计算自助样本 $\boldsymbol{X}^*$ 的样本中位数 M^* 和尺度统计值 Y^*，再计算 $T=(M^*-M)/Y^*$．多次重复该过程可得到来自于自助样本大量的 v 个 $T^{(1)},\cdots,T^{(v)}$，再令 $-d_2$ 和 d_1 为 $T^{(i)}$ 值的样本分位数（例如 $\alpha_0/2$ 与 $1-\alpha_0/2$）．

例 11.4.6 对中位数的 t-百分位数置信区间 考虑例 7.4.2 中奶酪中的 $n=10$ 个乳酸含量．通过 $v=10\ 000$ 次自助模拟得到乳酸含量中位数 θ 的 $1-\alpha_0=0.9$ 的置信区间．该样本中位数的值为 $M=1.41$，该中位数的绝对偏差为 $Y=0.245$．$(M^{*(i)}-M)/Y^{*(i)}$ 在 0.05 和 0.95 的样本分位数的值分别为 -2.133 和 1.581．这使得 t-百分位自助置信区间为 $(1.41-1.581\times0.245,1.41+2.133\times0.245)=(1.023,1.933)$．与之相比，$M^{*(i)}-M$ 值的 0.05 和 0.95 的样本分位数分别为 -0.32 和 0.16．这使得百分位自助区间等于 $(1.41-0.16,1.41+0.32)=(1.25,1.73)$．◄

在例 11.4.6 中，t-百分位数的区间要比百分数区间更宽．从而反映了一个事实，即根据自助样本得到的 Y^* 值的分散程度相当广．即暗示样本中我们期望看到的分散程度的变化较大．因此，认为 M^*-M 的分布的分散度与 $M-\theta$ 的分布的分散度相同的看法并不好．一般来说，在 t-百分位数和百分位数的自助区间都可以得到时，t-百分位数自助区间比百分位数自助区间更好．这其实是由于 $(M^*-M)/Y^*$ 的分布对 $\hat{F}$ 的依赖比 M^*-M 的分布对 $\hat{F}$ 的依赖要少．特别地，$(M^*-M)/Y^*$ 不依赖于分布 $\hat{F}$ 的任何尺度参数．由于该原因，我们更期望 $(M^*-M)/Y^*$ 的分布与 $(M-\theta)/Y$ 的相

似程度比 M^*-M 与 $M-\theta$ 的分布的相似程度更高.

例 11.4.7 样本相关系数的分布的特征 令 (X,Y) 有对二者而言方差均有限的二元联合分布 F，这样讨论相关系数就有意义了. 假设有来自于分布 F 的一组随机样本 $(X_1,Y_1),\cdots,(X_n,Y_n)$. 进一步假设我们对该样本的以下相关系数的分布感兴趣:

$$R=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\left(\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]\left[\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]\right)^{1/2}}.\tag{11.4.2}$$

我们可能对作为 X 和 Y 的相关系数 ρ 的估计 R 的方差、R 的偏差或者 R 一些其他特征感兴趣. 但无论目的是什么，都可使用非参数的方法. 例如，考虑 R 作为 ρ 的估计的偏差，该偏差等于 $\eta(X,F)=R-\rho$ 的均值. 我们从以观察到的成对数据的样本分布 F_n 来代替联合分布 F 开始. 该 F_n 是一个成对实数的离散型联合分布，它分配给 n 个观察样本的每一对的概率均为 $1/n$. 如果 (X^*,Y^*) 服从 F_n 分布，很容易就可以检验 X^* 和 Y^* 的相关系数为 R（见习题8）. 那么可以选择一个大的数 v，并从分布 F_n 中模拟出 v 个容量为 n 的样本. 对于每一个 i，可以将第 i 个自助样本代入等式(11.4.2)来计算该样本相关系数 $R^{(i)}$. 对于每一个 i，计算 $T^{(i)}=R^{(i)}-R$，并且用均值 $\frac{1}{v}\sum_{i=1}^{v}T^{(i)}$ 来估计 $R-\rho$ 的均值.

759

举一个数值例子，考虑例 5.10.1 中跳甲的数据. 样本相关系数为 $R=0.6401$. 抽取 $v=10\ 000$ 个样本容量 $n=31$ 的自助样本. 10 000 个自助样本中平均样本相关系数为 0.6354，它的模拟标准误差为 0.001，然后可以估计该样本相关系数的偏差为 $0.6354-0.6401=-0.0047$. ◀

11.4.4 自助参数法

例 11.4.8 变异系数偏差的修正 分布的变异系数是其标准偏差与均值之比（通常，人们只计算来自正的随机变量的分布的变异系数）. 如果认为我们的数据 $X_1,\cdots,X_n$ 是来自于参数为 μ 和 σ^2 的对数正态分布，那么变异系数为 $\theta=(e^{\sigma^2}-1)^{1/2}$. 该变异系数的极大似然估计是 $\hat{\theta}=(e^{\hat{\sigma}^2}-1)^{1/2}$，此处 $\hat{\sigma}$ 是 σ 的极大似然估计. 我们预期该变异系数的极大似然估计是一个有偏估计，因为它是非线性的. 计算偏差是一件困难的事，但是可以用自助参数法去估计该偏差. σ 的极大似然估计 $\hat{\sigma}$ 是 $\ln(X_1),\cdots,\ln(X_n)$ 的样本方差的平方根. μ 的极大似然估计 $\hat{\mu}$ 是 $\ln(X_1),\cdots,\ln(X_n)$ 的样本均值. 可以模拟大量的来自于参数为 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 的对数正态分布的容量为 n 的随机样本. 对于每一个 i，计算第 i 个自助样本的样本标准偏差 $\hat{\sigma}^{*(i)}$. 我们用值 $T^{(i)}=(e^{[\hat{\sigma}^{*(i)}]^2}-1)^{1/2}-\hat{\theta}$ 的样本均值来估计 $\hat{\theta}$ 的偏差.

举一个例子，考虑例 5.6.7 中所介绍的滚珠轴承失效次数的数据. 如果认为这些数据服从对数正态分布，那么参数的极大似然估计为 $\hat{\mu}=4.150$ 和 $\hat{\sigma}=0.5217$. θ

的极大似然估计为 $\hat{\theta}=0.5593$. 从对数正态分布中抽取 10 000 个样本容量为 23 的样本，并且计算其样本方差. 然而，有一个更加简单的方法去做该模拟. $[\hat{\sigma}^{*(i)}]^2$ 的分布是自由度为 22 的 χ^2 分布的随机变量乘以 $0.5217^2/23$ 后的分布. 因此，只要抽取 10 000 个来自自由度为 22 的 χ^2 分布的随机变量，然后每一个随机变量都乘以 $0.5217^2/23$，记第 i 个为$[\hat{\sigma}^{*(i)}]^2$. 这样，10 000 个 $T^{(i)}$ 的样本均值为 $-0.018\,25$，即为 $\hat{\theta}$ 的偏差的参数自助估计(其中模拟标准误差为 9.47×10^{-4}). 由于我们对偏差的估计为负值，这就意味着得到的 $\hat{\theta}$ 比 θ 小. 为了修正偏差，可以在原来的估计 $\hat{\theta}$ 上加上 0.018 25，从而可产生一个新的估计 $0.5593+0.01825=0.5776$. ◀

例 11.4.9 估计一个统计量的标准偏差 假设 $X_1,\cdots,X_n$ 是来自于均值为 μ 方差为 σ^2 的正态分布的随机样本，要研究的是具有此分布的随机变量取值至多为 c 的概率. 即我们对估计 $\theta=\Phi([c-\mu]/\sigma)$ 感兴趣. θ 的极大似然估计是 $\hat{\theta}=\Phi([c-\overline{X}]/\hat{\sigma})$. 很难计算 $\hat{\theta}$ 的标准偏差的显式解. 然而，可以从均值为 $\bar{x}$、方差为 $\hat{\sigma}^2$ 的正态分布中抽取很多（记为 v 个）自助样本. 对于第 i 个自助样本，计算样本均值 $\bar{x}^{*(i)}$，样本标准差 $\hat{\sigma}^{*(i)}$，最后计算 $\hat{\theta}^{*(i)}=\Phi([c-\bar{x}^{*(i)}]/\hat{\sigma}^{*(i)})$. 可以用

760

$$\overline{\theta}^{*}=\frac{1}{v}\sum_{i=1}^{v}\hat{\theta}^{*(i)}$$

来估计 $\hat{\theta}$ 的均值（正如例 11.4.8 中一样，这些也可以用来估计 $\hat{\theta}$ 的偏差）. 从而 $\hat{\theta}$ 的标准偏差可以用 $\hat{\theta}^{*(i)}$ 值的样本标准差

$$Z=\left(\frac{1}{v}\sum_{i=1}^{v}(\hat{\theta}^{*(i)}-\hat{\theta}^{*})^2\right)^{1/2}$$

来估计.

举个例子，可以用护理院的数据. 设有 $n=18$ 个观察值，我们关心的是 $\Phi([200-\mu]/\sigma)$. μ 和 σ 的极大似然估计为 $\hat{\mu}=182.17$ 和 $\hat{\sigma}=72.22$. $\hat{\theta}$ 的观察值是 $\Phi([200-182.17]/72.22)=0.5975$. 从均值为 182.17 和方差为 72.22^2 的正态分布中模拟出 10 000 个样本容量为 18 的样本. 对于第 i 个样本，计算出 $\hat{\theta}^{*(i)}$，$i=1,\cdots,10\,000$，从而得到平均值为 $\overline{\theta}^{*}=0.6020$，样本标准差为 $Z=0.097\,68$.

我们可以分两步来计算对于自助估计近似的模拟标准误差. 首先，使用例 11.1.10 的方法，得到 $\hat{\theta}^{*(i)}$ 的样本方差 Z^2 的模拟标准误差. 在我们的例子中，该值为 1.365×10^{-4}. 第二步 Delta 方法，像例 11.1.8 一样，得到 Z^2 的平方根的模拟标准误差. 在我们的例子中，第二步得到的值为 6.986×10^{-4}. ◀

例 11.4.10 当方差不等情况下比较均值 假设有两个样本 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 分别来自于两个可能不同的正态分布. 即 $X_1,\cdots,X_m$ 是来自于均值为 μ_1、方差为 $\sigma_1{}^2$ 的正态分布的独立同分布样本，而 $Y_1,\cdots,Y_n$ 是来自于均值为 μ_2、方差为 $\sigma_2{}^2$ 的正态分布的独立同分布样本. 在 8.4 节中，我们知道在假定 $k=\sigma_2^2/\sigma_1^2$ 已知时如何检验原

假设 H_0：$\mu_1=\mu_2$ 和备择假设 $H_1:\mu_1\neq\mu_2$. 然而，如果不假定比率 k 已知，则不存在标准的检验.

假设即使我们不知道 k，也选择使用通常的两样本 t 检验. 即当 $|U|>c$ 时拒绝 H_0，其中 U 是等式(8.4.3)所定义的统计量，c 是自由度为 $m+n-2$ 的 t 分布的 $1-\alpha_0/2$分位数. 如果 $k\neq1$，则该检验不一定有水平 α_0. 可以用参数自助法来试着计算该检验水平. 事实上，参数自助法可以帮助我们选择一个不同的临界值 c^* 来检验，从而至少可以估计出犯第一类错误的概率为 α_0.

761

举一个例子，再次使用例8.4.2中的数据. 两个分布的方差的极大似然估计分别为 $\hat{\sigma}_1{}^2=0.04$(对于 X 数据)，$\hat{\sigma}_2{}^2=0.022$(对于 Y 数据). 第一类错误的概率是当原假设(即 $\mu_1=\mu_2$)为真时拒绝给定的原假设的概率. 因此，必须在 X 和 Y 数据有相同均值的情况下模拟自助样本. 由于在计算 U 时 X 与 Y 数据的样本平均值是相减的，所以对于这两个样本我们选择什么样的共同均值无关紧要.

因此，参数自助法可以按照如下步骤进行. 首先，选择一个比较大的数 v，对于 $i=1,\cdots,v$，模拟($\overline{X}^{*(i)},\overline{Y}^{*(i)},S_X^{2*(i)},S_Y^{2*(i)}$)，其中这4个随机变量都独立，且有下面的分布：

- $\overline{X}^{*(i)}$服从均值为0，方差为 $\hat{\sigma}_1{}^2/m$ 的正态分布.
- $\overline{Y}^{*(i)}$服从均值为0，方差为 $\hat{\sigma}_2{}^2/n$ 的正态分布.
- $S_X^{2*(i)}$ 是自由度为 $m-1$ 的χ^2 分布的随机变量与 $\hat{\sigma}_1{}^2$ 的乘积.
- $S_Y^{2*(i)}$ 是自由度为 $n-1$ 的χ^2 分布的随机变量与 $\hat{\sigma}_2{}^2$ 的乘积.

然后对于每个 i 计算

$$U^{(i)}=\frac{(m+n-2)^{1/2}(\overline{X}^{*(i)}-\overline{Y}^{*(i)})}{\left(\frac{1}{m}+\frac{1}{n}\right)^{1/2}(S_X^{2*(i)}+S_Y^{2*(i)})^{1/2}}.$$

对于常用的两样本 t-检验中的第一类误差的概率的自助估计，我们的模拟近似值为在所有的模拟中满足 $|U^{(i)}|>c$ 的值所占的比例.

当 $v=10\,000$ 时，我们对于若干不同的 c 值用前面所描述过的分析. 取 c 为自由度等于16的 t 分布的 $1-\alpha_0/2$ 分位数，其中 $\alpha_0=j/1000$，$j=1,\cdots,999$. 图11-10显示了根据检验的名义水平为 α_0 得到犯第一类错误概率的自助估计的模拟近似图. 虽然自助估计值普遍稍微偏大了一点，但是两者的结论惊人得相似. 比方说，当 $\alpha_0=0.05$ 时，自助估计为0.065.

接下来，我们用自助分析来修正本例中的两样本 t-检验的水平. 为此，令 Z 等于模拟的 $|U^{(i)}|$值的样本 $1-\alpha_0$ 的分位数. 如果要做一个水平为 α_0 的检验，可以用 Z 值来代替两样本 t-检验的 c 值，从而当 $|U|>Z$ 时，拒绝原假设. 例如，取 $\alpha_0=0.05$，

t 分布的 0.975 的分位数是 2.12，而模拟出的 $Z=2.277$. Z 的模拟标准误差为 0.0089(基于将 10 000 个自助样本，分成 10 个子样本，每个子样本的容量为 1000). ◀

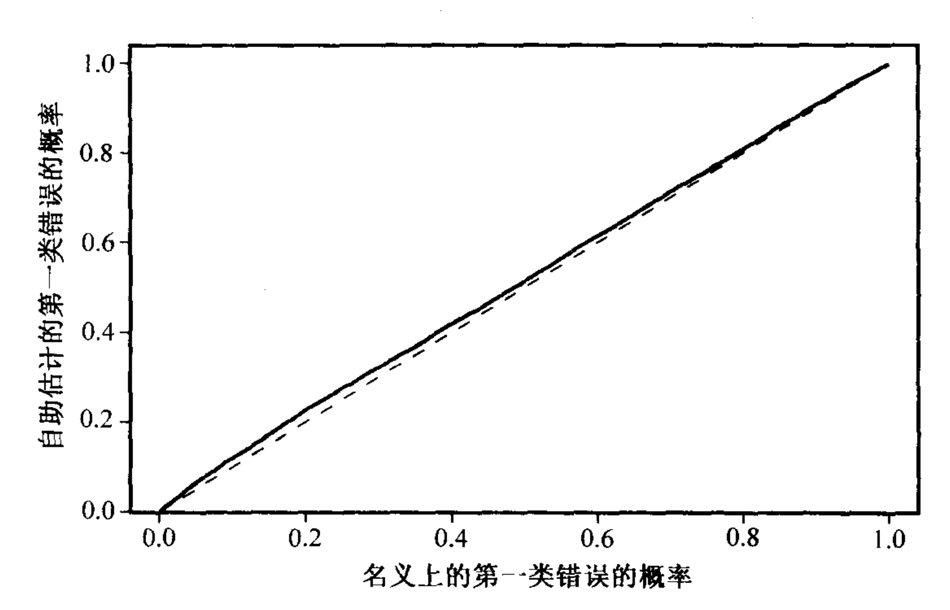


图 11-10 例 11.5.10 中 t-检验的自助估计的第一类错误概率相对于名义上第一类错误的概率的图形. 对角线上的虚线表示这两个错误概率相等

例 11.4.11 样本相关系数的偏差 在例 11.4.7 中，除了假定 X, Y 的方差有限外，对 (X,Y) 的分布 F 没有其他假设. 现在假设 (X,Y) 服从二元正态分布. 可以计算 6.4 节中习题 14 的所有参数的极大似然估计. 就如同例 11.2.6 一样，可以从参数为这些极大似然估计的二元正态分布模拟 v 个大小为 n 的样本. 对于样本 i, $i=1,\cdots,v$, 通过将第 i 个样本代到公式(11.4.2)中来计算样本相关系数 $R^{(i)}$. 对偏差的估计为 $\overline{R}-\hat{\rho}$. 注意：ρ 的极大似然估计 $\hat{\rho}$ 等于 R.

762

举一个数值的例子，考虑例 5.10.1 中的跳甲数据，其样本的相关系数为 $R=0.6401$. 构造一个服从正态分布、相关系数为 0.6401、样本容量 $n=31$ 的 $v=10\,000$ 个样本. 它的均值和方差不会影响 R 的分布(详见习题 12). 10 000 个自助样本的平均样本相关系数为 0.6352，模拟标准误差为 0.001，然后估计样本相关系数的偏差为 $0.6352-0.6401=-0.0049$. 而与我们在例 11.4.7 中用非参数自助法得到的值非常一致. ◀

11.4.5 小结

自助法是一种用来估计数据 X 和未知分布 F 的函数 η 的概率特征的方法. 即假设我们感兴趣的是 $\eta(X,F)$ 的均值、分位数或一些其他特征. 自助法的第一步就是用和已知分布 F 有些类似的分布 $\hat{F}$ 来代替 F. 然后，用 $\hat{F}$ 中的样本数据 X^* 代替 X. 最后，计算 $\eta(X^*,\hat{F})$ 的均值、分位数和其他特征，作为自助法估计. 除了在最简单的

情形，一般最后一步都是需要模拟. 基于 $\hat{F}$ 的不同选择，有两个不同的自助法. 在非参数自助法中，用样本分布函数作为 $\hat{F}$. 在参数自助法中，F 被假定为来自某个参数族，并且将分布 F 中未知的参数用它的 M. L. E. 或其他估计代替来得到 $\hat{F}$.

763

习题

1. 假设 $X_1,\cdots,X_n$ 是来自于服从参数为 θ 的指数分布的一组随机样本. 解释如何使用参数的自助法来估计样本均值 $\overline{X}$ 的方差（不需要模拟）.
2. 令 $x_1,\cdots,x_n$ 为一组随机样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的观察值. 令 F_n 为样本的分布函数. 令 $J_1,\cdots,J_n$ 为从一组数字 $\{1,\cdots,n\}$ 中有放回地抽取出的随机样本. 定义 $X_i^*=x_{J_i}$，$i=1,\cdots,n$. 证明：$\boldsymbol{X}^*=(X_1^*,\cdots,X_n^*)$ 是分布 F_n 的一个独立同分布的样本.
3. 令 n 为奇数，$\boldsymbol{X}=(X_1,\cdots,X_n)$ 是一个来自于某个分布、容量为 n 的样本. 假定希望使用非参数自助法来估计样本中位数的一些特性. 计算非参数自助样本的样本中位数是原始数据 $\boldsymbol{X}$ 中的最小的观测值的概率.
4. 使用在 10.2 节表格 10-5 中的第一列数据. 数据给出了 Forbes 实验中在 17 个不同的地方的水的沸点. 令 F 是选取这些沸点的分布. 我们不对 F 做太多假设. 假设对用样本中位数的偏差作为分布 F 中位数的偏差感兴趣. 使用非参数的自助法来估计该偏差. 首先，计算模拟近似的模拟标准差，然后确定为使你的偏差估计(分布为 $\hat{F}$)与正确的偏差(分布为 $\hat{F}$)之差在 0.02 内的概率至少为 0.9，需要多少自助样本.
5. 使用 9.2 节表格 9-4 中的数据. 若对使用样本中位数的偏差估计分布中位数的偏差感兴趣.
   a. 使用非参数自助法估计该偏差.
   b. 需要多少自助样本才使得估计的偏差在 0.05 以内的概率为 0.99?
6. 设一组数据为：49，－67.8，8，16，6，23，28，41，14，29，56，24，75，60，－48.
   a. 使用非参数自助法来估计该方差.
   b. 需要多少自助样本可使得方差估计的偏差以 0.95 的概率小于 0.05?
7. 使用在 8.4 节的习题 10 中描述过的表 8-1 中的血压数据，假定现在我们不确定两个治疗组的方差是否一样. 如同例 11.4.10 一样做一个非参数自助分析. 使用 $\nu=10\,000$ 次自助模拟.
   a. 对一个给定名义水平 $\alpha_0=0.1$ 的两样本的 t-检验，估计第一类错误发生的概率.
   b. 通过计算 $|U^i|$ 的自助分布的近似分位数来修正两样本 t-检验的水平.
   c. 计算 b. 中的分位数的模拟标准误差.
8. 在例 11.4.7 中，设样本分布 F_n 中随机抽取(X^*,Y^*)，证明：X^* 与 Y^* 之间的相关系数为(11.4.2)式中的 R.
9. 使用 10.2 节表格 10-6 中的鱼价的数据. 假定，我们仅假设 1970 和 1980 年鱼价的分布是一个方差有限的连续的联合分布. 我们对样本相关系数的特性感兴趣，构建 1000 个非参数自助样本来解此题.
   a. 用自助估计近似样本相关系数的方差.

b. 用自助估计近似样本相关系数的偏差.

c. 计算以上每个自助估计的模拟标准误差.

10. 使用7.5节中习题7的牛肉热狗数据. 模拟10 000个非参数自助样本来解此题.

a. 对牛肉热狗中卡路里的中位数，近似构造一个90%分位自助置信区间.

b. 对牛肉热狗中卡路里的中位数，近似构造一个90%的t-分位数自助法的置信区间.

c. 比较这些区间与假定数据来自正态分布所形成的90%的区间.

11. 在11.1节中习题14定义过随机变量的偏度. 假定$X_1,\cdots,X_n$是来自于分布为F的一组随机样本. 样本的偏度被定义为

$$M_3=\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^{3/2}}.$$

764

可以用M_3作为分布F的偏度θ的一个估计. 自助法可以用来估计样本偏度作为θ估计值的偏差和标准差.

a. 证明：M_3是样本分布F_n的偏度.

b. 使用10.2节表10-6中的1970年的鱼价的数据，计算样本偏度，并且模拟1000个自助样本，使用这些自助样本来估计样本偏度的偏差和标准差.

12. 假设$(X_1,Y_1),\cdots,(X_n,Y_n)$来自于均值分别为$\mu_x$与$\mu_y$、方差分别为$\sigma_x^2$和$\sigma_y^2$、相关系数为$\rho$的二元正态分布的一组随机样本. 令$R$为样本的相关系数. 证明：$R$的分布只与$\rho$有关，而与$\mu_x$、$\mu_y$、$\sigma_x^2$和$\sigma_y^2$无关.

附　　表

二项分布表

$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

n	k	p = 0.1	p = 0.2	p = 0.3	p = 0.4	p = 0.5
2	0	.8100	.6400	.4900	.3600	.2500
	1	.1800	.3200	.4200	.4800	.5000
	2	.0100	.0400	.0900	.1600	.2500
3	0	.7290	.5120	.3430	.2160	.1250
	1	.2430	.3840	.4410	.4320	.3750
	2	.0270	.0960	.1890	.2880	.3750
	3	.0010	.0080	.0270	.0640	.1250
4	0	.6561	.4096	.2401	.1296	.0625
	1	.2916	.4096	.4116	.3456	.2500
	2	.0486	.1536	.2646	.3456	.3750
	3	.0036	.0256	.0756	.1536	.2500
	4	.0001	.0016	.0081	.0256	.0625
5	0	.5905	.3277	.1681	.0778	.0312
	1	.3280	.4096	.3602	.2592	.1562
	2	.0729	.2048	.3087	.3456	.3125
	3	.0081	.0512	.1323	.2304	.3125
	4	.0005	.0064	.0284	.0768	.1562
	5	.0000	.0003	.0024	.0102	.0312
6	0	.5314	.2621	.1176	.0467	.0156
	1	.3543	.3932	.3025	.1866	.0938
	2	.0984	.2458	.3241	.3110	.2344
	3	.0146	.0819	.1852	.2765	.3125
	4	.0012	.0154	.0595	.1382	.2344
	5	.0001	.0015	.0102	.0369	.0938
	6	.0000	.0001	.0007	.0041	.0156
7	0	.4783	.2097	.0824	.0280	.0078
	1	.3720	.3670	.2471	.1306	.0547
	2	.1240	.2753	.3176	.2613	.1641
	3	.0230	.1147	.2269	.2903	.2734
	4	.0026	.0287	.0972	.1935	.2734
	5	.0002	.0043	.0250	.0774	.1641
	6	.0000	.0004	.0036	.0172	.0547
	7	.0000	.0000	.0002	.0016	.0078

（续）

n	k	$p=0.1$	$p=0.2$	$p=0.3$	$p=0.4$	$p=0.5$
8	0	.4305	.1678	.0576	.0168	.0039
	1	.3826	.3355	.1977	.0896	.0312
	2	.1488	.2936	.2965	.2090	.1094
	3	.0331	.1468	.2541	.2787	.2188
	4	.0046	.0459	.1361	.2322	.2734
	5	.0004	.0092	.0467	.1239	.2188
	6	.0000	.0011	.0100	.0413	.1094
	7	.0000	.0001	.0012	.0079	.0312
	8	.0000	.0000	.0001	.0007	.0039
9	0	.3874	.1342	.0404	.0101	.0020
	1	.3874	.3020	.1556	.0605	.0176
	2	.1722	.3020	.2668	.1612	.0703
	3	.0446	.1762	.2668	.2508	.1641
	4	.0074	.0661	.1715	.2508	.2461
	5	.0008	.0165	.0735	.1672	.2461
	6	.0001	.0028	.0210	.0743	.1641
	7	.0000	.0003	.0039	.0212	.0703
	8	.0000	.0000	.0004	.0035	.0176
	9	.0000	.0000	.0000	.0003	.0020
10	0	.3487	.1074	.0282	.0060	.0010
	1	.3874	.2684	.1211	.0403	.0098
	2	.1937	.3020	.2335	.1209	.0439
	3	.0574	.2013	.2668	.2150	.1172
	4	.0112	.0881	.2001	.2508	.2051
	5	.0015	.0264	.1029	.2007	.2461
	6	.0001	.0055	.0368	.1115	.2051
	7	.0000	.0008	.0090	.0425	.1172
	8	.0000	.0001	.0014	.0106	.0439
	9	.0000	.0000	.0001	.0016	.0098
	10	.0000	.0000	.0000	.0001	.0010

（续）

n	k	$p=0.1$	$p=0.2$	$p=0.3$	$p=0.4$	$p=0.5$
15	0	.2059	.0352	.0047	.0005	.0000
	1	.3432	.1319	.0305	.0047	.0005
	2	.2669	.2309	.0916	.0219	.0032
	3	.1285	.2501	.1700	.0634	.0139
	4	.0428	.1876	.2186	.1268	.0417
	5	.0105	.1032	.2061	.1859	.0916
	6	.0019	.0430	.1472	.2066	.1527
	7	.0003	.0138	.0811	.1771	.1964
	8	.0000	.0035	.0348	.1181	.1964
	9	.0000	.0007	.0116	.0612	.1527
	10	.0000	.0001	.0030	.0245	.0916
	11	.0000	.0000	.0006	.0074	.0417
	12	.0000	.0000	.0001	.0016	.0139
	13	.0000	.0000	.0000	.0003	.0032
	14	.0000	.0000	.0000	.0000	.0005
	15	.0000	.0000	.0000	.0000	.0000
20	0	.1216	.0115	.0008	.0000	.0000
	1	.2701	.0576	.0068	.0005	.0000
	2	.2852	.1369	.0278	.0031	.0002
	3	.1901	.2054	.0716	.0123	.0011
	4	.0898	.2182	.1304	.0350	.0046
	5	.0319	.1746	.1789	.0746	.0148
	6	.0089	.1091	.1916	.1244	.0370
	7	.0020	.0545	.1643	.1659	.0739
	8	.0003	.0222	.1144	.1797	.1201
	9	.0001	.0074	.0654	.1597	.1602
	10	.0000	.0020	.0308	.1171	.1762
	11	.0000	.0005	.0120	.0710	.1602
	12	.0000	.0001	.0039	.0355	.1201
	13	.0000	.0000	.0010	.0146	.0739
	14	.0000	.0000	.0002	.0049	.0370
	15	.0000	.0000	.0000	.0013	.0148
	16	.0000	.0000	.0000	.0003	.0046
	17	.0000	.0000	.0000	.0000	.0011
	18	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000

泊松分布表

$\Pr(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$

k	$\lambda=.1$	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0003	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

k	$\lambda=1.5$	2	3	4	5	6	7	8	9	10
0	.2231	.1353	.0498	.0183	.0067	.0025	.0009	.0003	.0001	.0000
1	.3347	.2707	.1494	.0733	.0337	.0149	.0064	.0027	.0011	.0005
2	.2510	.2707	.2240	.1465	.0842	.0446	.0223	.0107	.0050	.0023
3	.1255	.1804	.2240	.1954	.1404	.0892	.0521	.0286	.0150	.0076
4	.0471	.0902	.1680	.1954	.1755	.1339	.0912	.0573	.0337	.0189
5	.0141	.0361	.1008	.1563	.1755	.1606	.1277	.0916	.0607	.0378
6	.0035	.0120	.0504	.1042	.1462	.1606	.1490	.1221	.0911	.0631
7	.0008	.0034	.0216	.0595	.1044	.1377	.1490	.1396	.1171	.0901
8	.0001	.0009	.0081	.0298	.0653	.1033	.1304	.1396	.1318	.1126
9	.0000	.0002	.0027	.0132	.0363	.0688	.1014	.1241	.1318	.1251
10	.0000	.0000	.0008	.0053	.0181	.0413	.0710	.0993	.1186	.1251
11	.0000	.0000	.0002	.0019	.0082	.0225	.0452	.0722	.0970	.1137
12	.0000	.0000	.0001	.0006	.0034	.0113	.0264	.0481	.0728	.0948
13	.0000	.0000	.0000	.0002	.0013	.0052	.0142	.0296	.0504	.0729
14	.0000	.0000	.0000	.0001	.0005	.0022	.0071	.0169	.0324	.0521
15	.0000	.0000	.0000	.0000	.0002	.0009	.0033	.0090	.0194	.0347
16	.0000	.0000	.0000	.0000	.0000	.0003	.0014	.0045	.0109	.0217
17	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0021	.0058	.0128
18	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0029	.0071
19	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0037
20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0006	.0019
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0009
22	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004
23	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

χ^2 分布表

如果随机变量 X 具有自由度为 n 的 χ^2 分布，下表给出满足 $\Pr(X \leqslant x) = p$ 的 x 的值，即 X 的 p 分位数.

	p								
n	.005	.01	.025	.05	.10	.20	.25	.30	.40
1	.0000	.0002	.0010	.0039	.0158	.0642	.1015	.1484	.2750
2	.0100	.0201	.0506	.1026	.2107	.4463	.5754	.7133	1.022
3	.0717	.1148	.2158	.3518	.5844	1.005	1.213	1.424	1.869
4	.2070	.2971	.4844	.7107	1.064	1.649	1.923	2.195	2.753
5	.4117	.5543	.8312	1.145	1.610	2.343	2.675	3.000	3.655
6	.6757	.8721	1.237	1.635	2.204	3.070	3.455	3.828	4.570
7	.9893	1.239	1.690	2.167	2.833	3.822	4.255	4.671	5.493
8	1.344	1.647	2.180	2.732	3.490	4.594	5.071	5.527	6.423
9	1.735	2.088	2.700	3.325	4.168	5.380	5.899	6.393	7.357
10	2.156	2.558	3.247	3.940	4.865	6.179	6.737	7.267	8.295
11	2.603	3.053	3.816	4.575	5.578	6.989	7.584	8.148	9.237
12	3.074	3.571	4.404	5.226	6.304	7.807	8.438	9.034	10.18
13	3.565	4.107	5.009	5.892	7.042	8.634	9.299	9.926	11.13
14	4.075	4.660	5.629	6.571	7.790	9.467	10.17	10.82	12.08
15	4.601	5.229	6.262	7.261	8.547	10.31	11.04	11.72	13.03
16	5.142	5.812	6.908	7.962	9.312	11.15	11.91	12.62	13.98
17	5.697	6.408	7.564	8.672	10.09	12.00	12.79	13.53	14.94
18	6.265	7.015	8.231	9.390	10.86	12.86	13.68	14.43	15.89
19	6.844	7.633	8.907	10.12	11.65	13.72	14.56	15.35	16.85
20	7.434	8.260	9.591	10.85	12.44	14.58	15.45	16.27	17.81
21	8.034	8.897	10.28	11.59	13.24	15.44	16.34	17.18	18.77
22	8.643	9.542	10.98	12.34	14.04	16.31	17.24	18.10	19.73
23	9.260	10.20	11.69	13.09	14.85	17.19	18.14	19.02	20.69
24	9.886	10.86	12.40	13.85	15.66	18.06	19.04	19.94	21.65
25	10.52	11.52	13.12	14.61	16.47	18.94	19.94	20.87	22.62
30	13.79	14.95	16.79	18.49	20.60	23.36	24.48	25.51	27.44
40	20.71	22.16	24.43	26.51	29.05	32.34	33.66	34.87	36.16
50	27.99	29.71	32.36	34.76	37.69	41.45	42.94	44.31	46.86
60	35.53	37.48	40.48	43.19	46.46	50.64	52.29	53.81	56.62
70	43.27	45.44	48.76	51.74	55.33	59.90	61.70	63.35	66.40
80	51.17	53.54	57.15	60.39	64.28	69.21	71.14	72.92	76.19
90	59.20	61.75	65.65	69.13	73.29	78.56	80.62	82.51	85.99
100	67.33	70.06	74.22	77.93	82.86	87.95	90.13	92.13	95.81

(续)

.50	.60	.70	.75	.80	.90	.95	.975	.99	.995
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.635	7.879
1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	9.210	10.60
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.34	12.84
3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.14	13.28	14.86
4.351	5.132	6.064	6.626	7.289	9.236	11.07	12.83	15.09	16.75
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.81	18.55
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.48	20.28
7.344	8.351	9.524	10.22	11.03	13.36	15.51	17.53	20.09	21.95
8.343	9.414	10.66	11.39	12.24	14.68	16.92	19.02	21.67	23.59
9.342	10.47	11.78	12.55	13.44	15.99	18.31	20.48	23.21	25.19
10.34	11.53	12.90	13.70	14.63	17.27	19.68	21.92	24.72	26.76
11.34	12.58	14.01	14.85	15.81	18.55	21.03	23.34	26.22	28.30
12.34	13.64	15.12	15.98	16.98	19.81	22.36	24.74	27.69	29.82
13.34	14.69	16.22	17.12	18.15	21.06	23.68	26.12	29.14	31.32
14.34	15.73	17.32	18.25	19.31	22.31	25.00	27.49	30.58	32.80
15.34	16.78	18.42	19.37	20.47	23.54	26.30	28.85	32.00	34.27
16.34	17.82	19.51	20.49	21.61	24.77	27.59	30.19	33.41	35.72
17.34	18.87	20.60	21.60	22.76	25.99	28.87	31.53	34.81	37.16
18.34	19.91	21.69	22.72	23.90	27.20	30.14	32.85	36.19	38.58
19.34	20.95	22.77	23.83	25.04	28.41	31.41	34.17	37.57	40.00
20.34	21.99	23.86	24.93	26.17	29.62	32.67	35.48	38.93	41.40
21.34	23.03	24.94	26.04	27.30	30.81	33.92	36.78	40.29	42.80
22.34	24.07	26.02	27.14	28.43	32.01	35.17	38.08	41.64	44.18
23.34	25.11	27.10	28.24	29.55	33.20	36.42	39.36	42.98	45.56
24.34	26.14	28.17	29.34	30.68	34.38	37.65	40.65	44.31	46.93
29.34	31.32	33.53	34.80	36.25	40.26	43.77	46.98	50.89	53.67
39.34	41.62	44.16	45.62	47.27	51.81	55.76	59.34	63.69	66.77
49.33	51.89	54.72	56.33	58.16	63.17	67.51	71.42	76.15	79.49
59.33	62.13	65.23	66.98	68.97	74.40	79.08	83.30	88.38	91.95
69.33	72.36	75.69	77.58	79.71	85.53	90.53	95.02	100.4	104.2
79.33	82.57	86.12	88.13	90.41	96.58	101.9	106.6	112.3	116.3
89.33	92.76	96.52	98.65	101.1	107.6	113.1	118.1	124.1	128.3
99.33	102.9	106.9	109.1	111.7	118.5	124.3	129.6	135.8	140.2

t 分布表

如果随机变量 X 具有自由度为 n 的 t 分布，下表给出满足 $\Pr(X \leqslant x)=p$ 的 x 的值．

n	$p=.55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845

（续）

n	$p=.55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576

标准正态分布表

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{(2\pi)^{1/2}}\exp\left(-\frac{1}{2}u^{2}\right)\mathrm{d}u$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.62	0.9956
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.64	0.9959
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.66	0.9961
0.19	0.5753	0.79	0.7852	1.39	0.9177	1.99	0.9767	2.68	0.9963
0.20	0.5793	0.80	0.7881	1.40	0.9192	2.00	0.9773	2.70	0.9965
0.21	0.5832	0.81	0.7910	1.41	0.9207	2.01	0.9778	2.72	0.9967
0.22	0.5871	0.82	0.7939	1.42	0.9222	2.02	0.9783	2.74	0.9969
0.23	0.5910	0.83	0.7967	1.43	0.9236	2.03	0.9788	2.76	0.9971
0.24	0.5948	0.84	0.7995	1.44	0.9251	2.04	0.9793	2.78	0.9973
0.25	0.5987	0.85	0.8023	1.45	0.9265	2.05	0.9798	2.80	0.9974
0.26	0.6026	0.86	0.8051	1.46	0.9279	2.06	0.9803	2.82	0.9976
0.27	0.6064	0.87	0.8079	1.47	0.9292	2.07	0.9808	2.84	0.9977
0.28	0.6103	0.88	0.8106	1.48	0.9306	2.08	0.9812	2.86	0.9979
0.29	0.6141	0.89	0.8133	1.49	0.9319	2.09	0.9817	2.88	0.9980
0.30	0.6179	0.90	0.8159	1.50	0.9332	2.10	0.9821	2.90	0.9981
0.31	0.6217	0.91	0.8186	1.51	0.9345	2.11	0.9826	2.92	0.9983
0.32	0.6255	0.92	0.8212	1.52	0.9357	2.12	0.9830	2.94	0.9984
0.33	0.6293	0.93	0.8238	1.53	0.9370	2.13	0.9834	2.96	0.9985
0.34	0.6331	0.94	0.8264	1.54	0.9382	2.14	0.9838	2.98	0.9986
0.35	0.6368	0.95	0.8289	1.55	0.9394	2.15	0.9842	3.00	0.9987
0.36	0.6406	0.96	0.8315	1.56	0.9406	2.16	0.9846	3.05	0.9989
0.37	0.6443	0.97	0.8340	1.57	0.9418	2.17	0.9850	3.10	0.9990
0.38	0.6480	0.98	0.8365	1.58	0.9429	2.18	0.9854	3.15	0.9992
0.39	0.6517	0.99	0.8389	1.59	0.9441	2.19	0.9857	3.20	0.9993
0.40	0.6554	1.00	0.8413	1.60	0.9452	2.20	0.9861	3.25	0.9994
0.41	0.6591	1.01	0.8437	1.61	0.9463	2.21	0.9864	3.30	0.9995
0.42	0.6628	1.02	0.8461	1.62	0.9474	2.22	0.9868	3.35	0.9996
0.43	0.6664	1.03	0.8485	1.63	0.9485	2.23	0.9871	3.40	0.9997
0.44	0.6700	1.04	0.8508	1.64	0.9495	2.24	0.9875	3.45	0.9997
0.45	0.6736	1.05	0.8531	1.65	0.9505	2.25	0.9878	3.50	0.9998
0.46	0.6772	1.06	0.8554	1.66	0.9515	2.26	0.9881	3.55	0.9998
0.47	0.6808	1.07	0.8577	1.67	0.9525	2.27	0.9884	3.60	0.9998
0.48	0.6844	1.08	0.8599	1.68	0.9535	2.28	0.9887	3.65	0.9999
0.49	0.6879	1.09	0.8621	1.69	0.9545	2.29	0.9890	3.70	0.9999
0.50	0.6915	1.10	0.8643	1.70	0.9554	2.30	0.9893	3.75	0.9999
0.51	0.6950	1.11	0.8665	1.71	0.9564	2.31	0.9896	3.80	0.9999
0.52	0.6985	1.12	0.8686	1.72	0.9573	2.32	0.9898	3.85	0.9999
0.53	0.7019	1.13	0.8708	1.73	0.9582	2.33	0.9901	3.90	1.0000
0.54	0.7054	1.14	0.8729	1.74	0.9591	2.34	0.9904	3.95	1.0000
0.55	0.7088	1.15	0.8749	1.75	0.9599	2.35	0.9906	4.00	1.0000
0.56	0.7123	1.16	0.8770	1.76	0.9608	2.36	0.9909		
0.57	0.7157	1.17	0.8790	1.77	0.9616	2.37	0.9911		
0.58	0.7190	1.18	0.8810	1.78	0.9625	2.38	0.9913		
0.59	0.7224	1.19	0.8830	1.79	0.9633	2.39	0.9916		

F 分布的 0.95 分位数表

如果随机变量 X 具有自由度为 m，n 的 F 分布，下表给出满足 $\Pr(X \leqslant x)=0.95$ 的 x 的值．

n \ m	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	245.9	248.0	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.62	2.58	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.16	2.11	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.95	1.90	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.75	1.66	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.32	1.22	1.00

F 分布的 0.975 分位数表

如果随机变量 X 具有自由度为 m，n 的 F 分布，下表给出满足 $\Pr(X \leqslant x)=0.975$ 的 x 的值.

n \ m	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	984.9	993.1	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.43	39.45	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.25	14.17	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.26	3.20	3.14	3.08
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.59	2.52	2.46	2.40
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.29	2.22	2.16	2.09
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.94	1.82	1.69	1.61	1.58	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.48	1.39	1.27	1.00